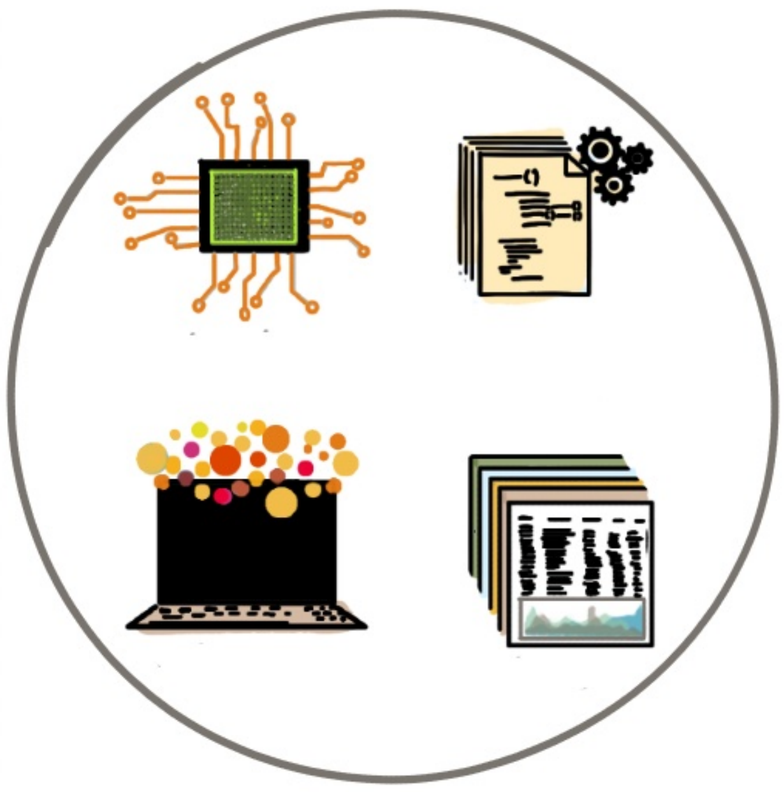


AUTONOMY

By

Aidanjali Venkatiaman

INTRODUCTION

AUTONOMY IS THE ASPIRATION THAT MACHINES THINK AND ACT INDEPENDENTLY, MAKING DECISIONS IN THE PLACE OF HUMANS.

WE ARE ALREADY SEEING DRONES AND SELF-DRIVING CARS ACHIEVING SOME OF THESE GOALS.

AUTONOMOUS SYSTEMS CAN EXTEND HUMAN CAPABILITIES USING ARTIFICIAL INTELLIGENCE. THEY ARE SUBJECT TO THE SAME HYPE AND FEAR ASSOCIATED WITH AI. SO THE PROSPECT OF ROUTINELY ENCOUNTERING SUCH MACHINES IN DAILY LIFE IS AT ONCE FASCINATING AND TERRIFYING.

WHAT FEATURES AND USES DO THEY HAVE? HOW DO WE KNOW THEY ARE SAFE AND TRUSTWORTHY? ARE WE READY AND PREPARED FOR LARGE-SCALE ADOPTION?

THIS BOOK EXPLORES THE CONCEPT OF AUTONOMY, HOW IT CAN POSITIVELY IMPACT HUMAN LIVES, HOW IT CAN BE INTRODUCED WITH THE RIGHT PROCESSES & SUPPORT SYSTEMS. WE ALSO EXAMINE ITS LIMITATIONS & ASSOCIATED ETHICAL ISSUES.

AN ILLUSTRATED GUIDE TO ARTIFICIAL INTELLIGENCE IN THIS SERIES IS A RECOMMENDED PRE-READ.

SCOPE

WHAT IS

- AUTONOMY
- AUTOMATION
- ROBOT
- 'DEGREES' OF AUTONOMY

A CLOSER LOOK

- DESCRIPTION
- FEATURES / ABILITIES
- USES IN SAFETY/WAR

BEHIND THE SCENES

- ARTIFICIAL INTELLIGENCE
- DATA
- TESTING
- ENGINEERING
- CHALLENGES

IS IT SAFE?

- STANDARDS
- REGULATION
- LAW

WHO DO WE TRUST?

- PEOPLE OR MACHINES?
- THE FOUR 'P's

WHO DO WE BLAME?

- HUMAN IN THE LOOP
- MORAL CRUMPLE ZONE

WHY DO WE CARE?

- ETHICS/DANGERS
- A BRIEF OUTLOOK

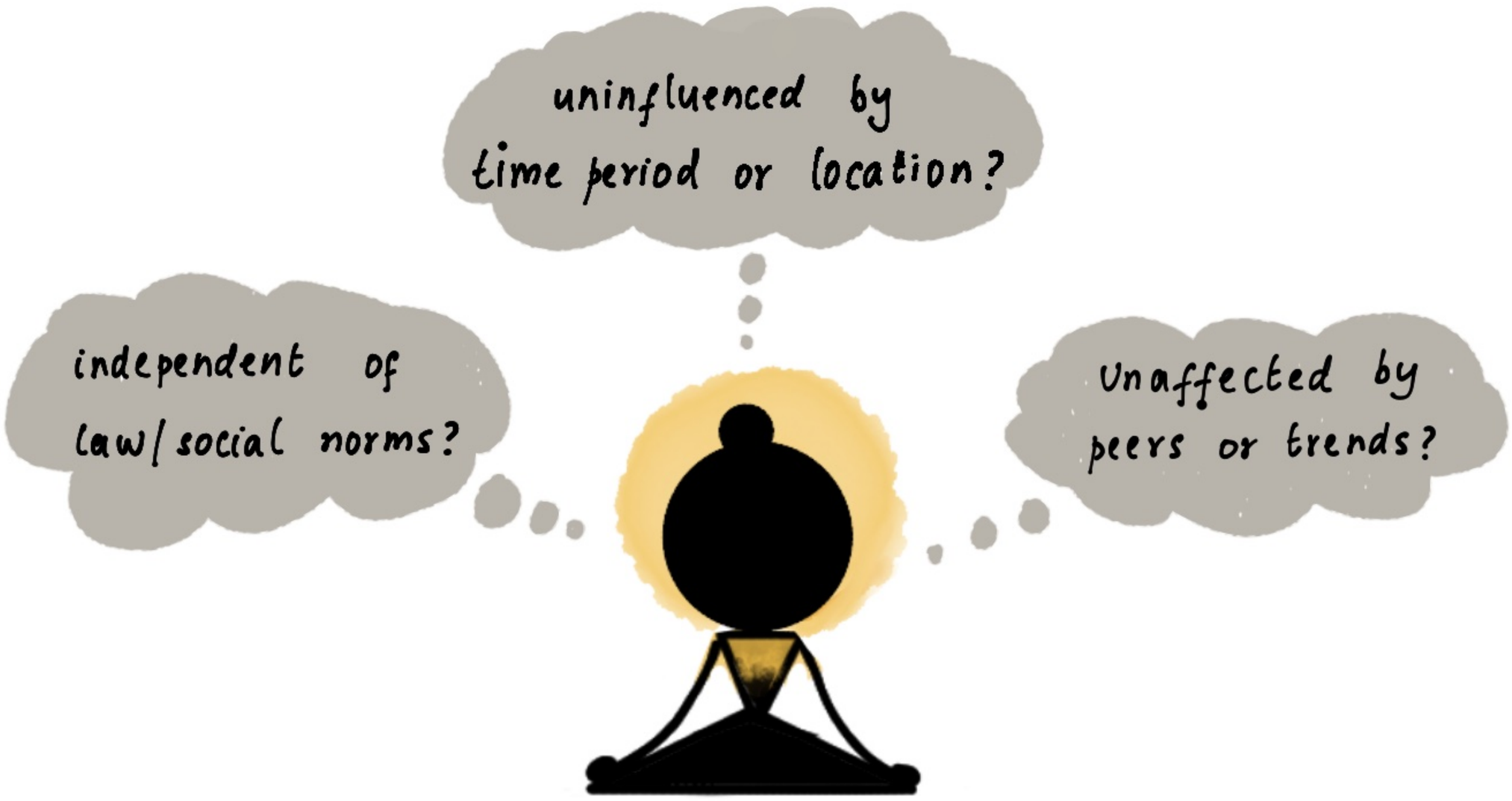
MORE TO EXPLORE

REFERENCES

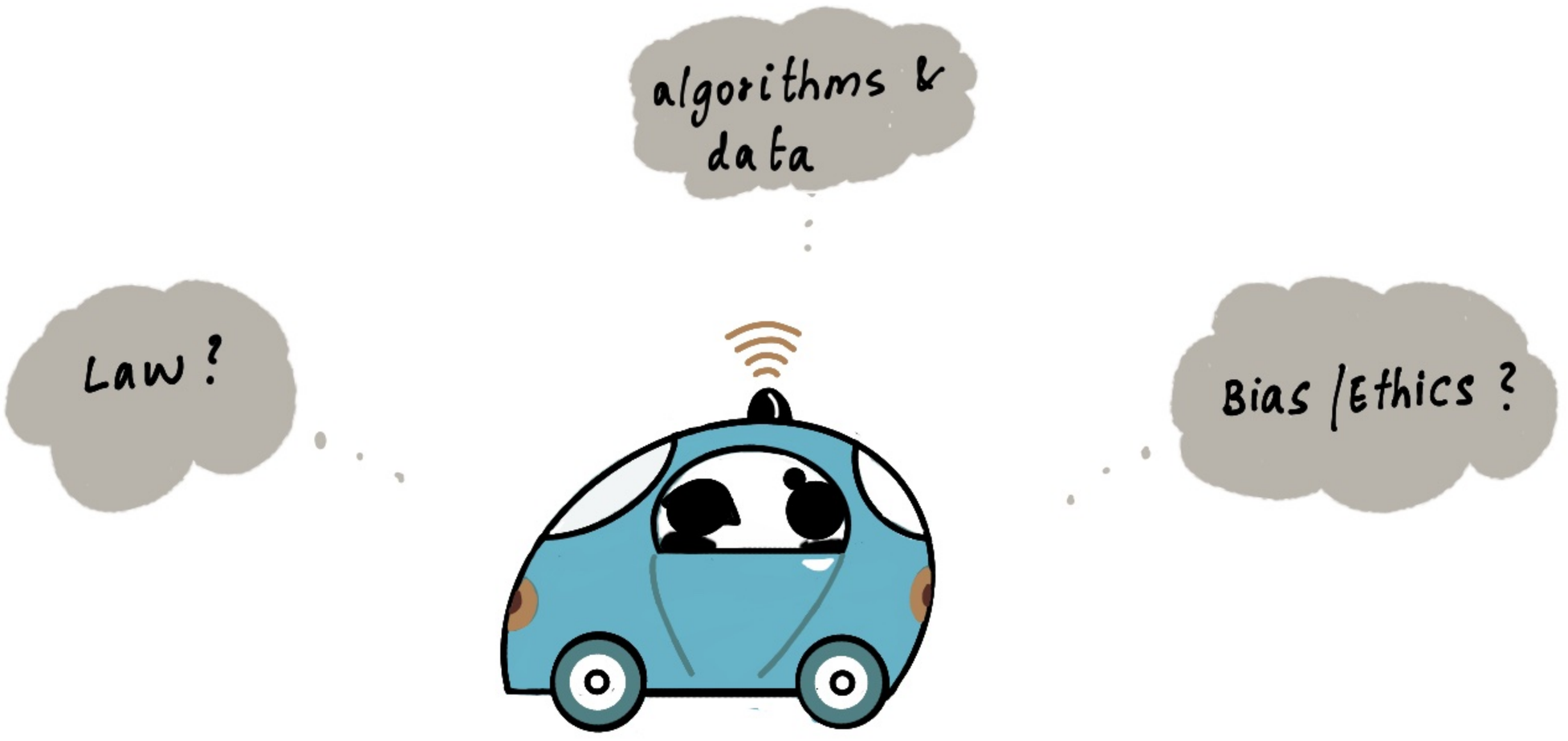
WHAT IS AUTONOMY?

AUTONOMY

AUTONOMY IS REGARDED AS A STATE OF BEING ABLE TO EXIST/ACT INDEPENTLY, SELF-DIRECT AND SELF-GOVERN



THIS IS AN AMBIGUOUS DEFINITION WHETHER APPLIED TO HUMANS OR TO THE MACHINES WE BUILD.



AN AUTONOMOUS SYSTEM CAN MAKE DECISIONS AND OPTIMISATIONS IN A CHANGING ENVIRONMENT WITHOUT (MUCH) HUMAN INSTRUCTION

SOME TERMS

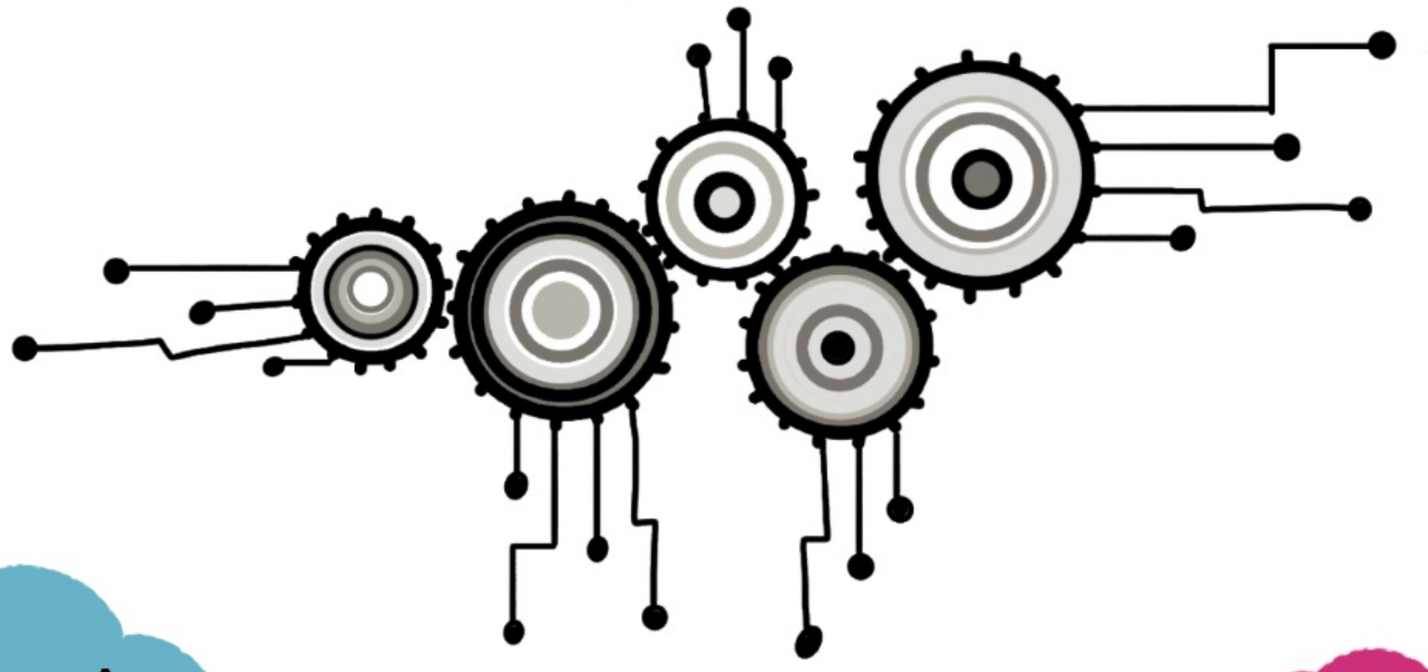
LET US START BY EXPLORING A FEW BASIC TERMS. AUTOMATION, ROBOTS AND 'DEGREES' OF AUTONOMY.

AUTOMATION

AUTOMATION

MECHANISING A TASK SO A HUMAN HAS VERY LITTLE TO DO

THIS MECHANISATION COULD BE THROUGH HARDWARE OR SOFTWARE.



Well defined algorithm

Predictable result

Same behaviour in all environments

Repeatable set of actions

Speeds up feedback loops

Preset assumptions

Cannot handle unplanned scenarios

THE AUTOMATION PARADOX IS THAT MORE AUTOMATION

- CAN LEAD TO NEW, UNFORESEEN ERRORS
- MAKES HUMAN INPUT EVEN MORE CRITICAL TO THE NON-AUTOMATED ASPECTS OF A TASK

AUTOMATION Vs AUTONOMY

TASK

CLEAN A FLOOR

AUTOMATION

USE A MACHINE TO HELP

AUTONOMY

GET THE MACHINE TO DO THE JOB

ROBOTS

ROBOT

A PROGRAMMABLE PHYSICAL DEVICE THAT CAN DO A SET OF TASKS

REPETITIVE

DIFFICULT

COGNITIVE

ROBOTS ARE NOT INTENDED TO REPLACE HUMANS ENTIRELY

ROBOTS ARE NOW ABLE TO ASSIST IN WELL-DEFINED TASKS AND IN TASKS THAT ARE MORE COMPLEX AND NEED COGNITIVE SKILLS.

A PERFECT ROBOT MAY EXIST ONLY AS A FICTIONAL WORK

LEONARDO DA VINCI'S HUMANOID 1400s

BICENTENNIAL MAN FILM 1990s

EMOTION READING ROBOT - 2014

ELSIE

ELMER

THE FIRST AUTONOMOUS ROBOTS WERE ELMER AND ELSIE, CREATED IN THE 1940s. THEY RESPONDED TO LIGHT AND TOUCH AND COULD TAKE THEMSELVES TO RECHARGE.

'DEGREES' OF AUTONOMY

AS IT STANDS, AUTONOMY IS NOT YET A BINARY CONCEPT. THERE ARE VARYING 'DEGREES' OF AUTONOMY.

MANUAL CONTROL / TELE OPERATION

- THE MACHINE EXERCISES NO INDEPENDENCE
- IT IS FULLY OPERATED AND CONTROLLED BY A HUMAN

SHARED AUTONOMY

- THE MACHINE WORKS CLOSELY WITH THE OPERATOR TAKING ON PART OF THE TASK
- THE OPERATOR STAYS ENGAGED

PERSISTENT AUTONOMY

THE MACHINE

- OPERATES
- ADAPTS
- SELF-HEALS

IN AN UNPREDICTABLE SETTING TO COMPLETE A TASK SET BY AN OPERATOR

FOR SELF-DRIVING CARS, THE SOCIETY OF AUTOMOTIVE ENGINEERS HAVE DEFINED SIX LEVELS OF AUTONOMY. sae.org/blog/sae-j3016-update

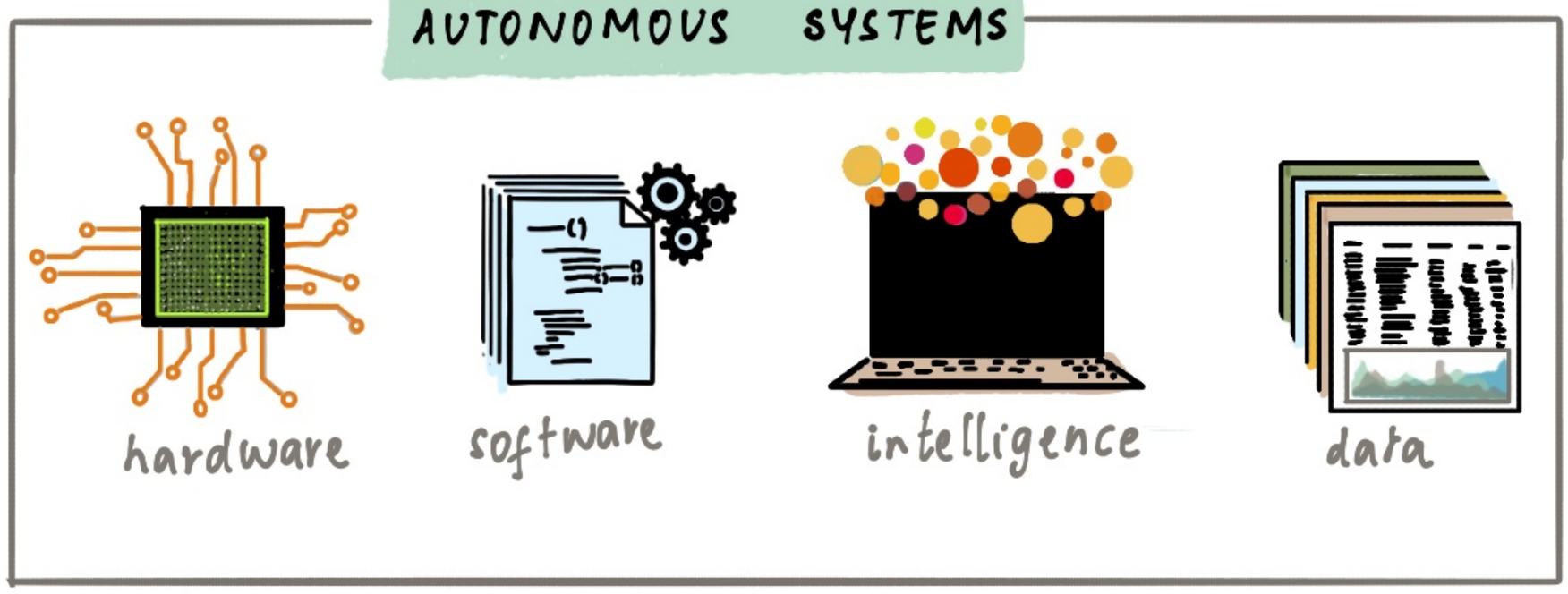
A CLOSER LOOK

IN THIS SECTION, WE DESCRIBE AUTONOMOUS SYSTEMS, FEATURES AND APPLICATIONS, AND WHY CONDITIONS ARE RIGHT FOR RESEARCH & DEVELOPMENT OF AUTONOMOUS SYSTEMS.

AUTONOMOUS SYSTEMS

ROBOTS AND AUTONOMOUS SYSTEMS ARE THE ARMS, LEGS AND SENSORS OF BIG DATA WORKING IN THE INTERNET OF THINGS.

-PROF DAVID LANE

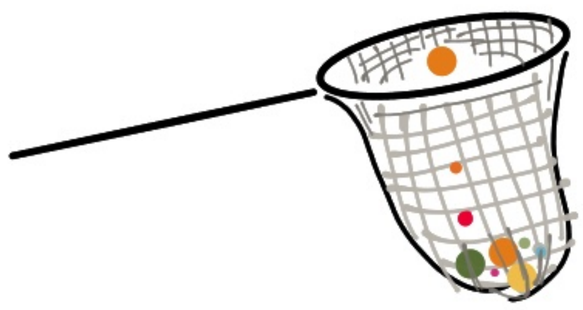


1

SENSE THE ENVIRONMENT

2

GATHER INFO ABOUT SURROUNDINGS

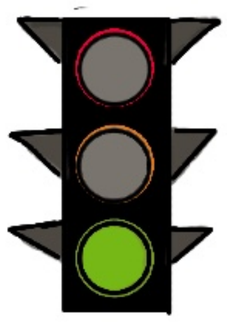


3

PLAN OUT WHAT NEXT

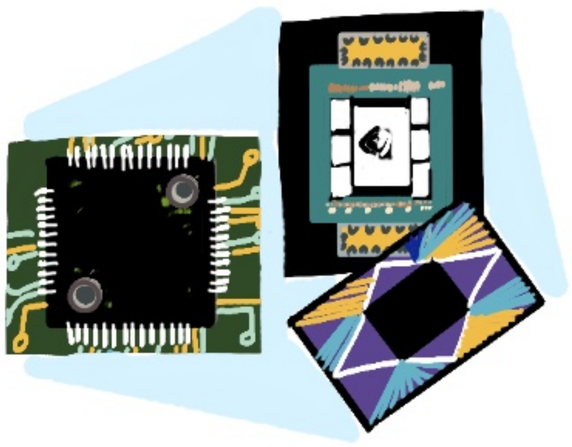
4

ACT WHEN IT IS SAFE



THE RIGHT CONDITIONS

HARDWARE

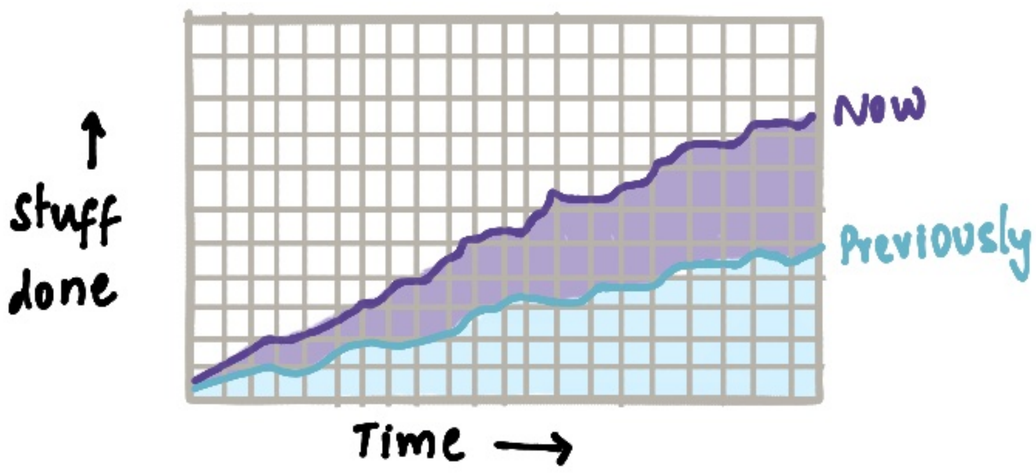


INCREASED COMPUTE AND STORAGE AVAILABLE

SOFTWARE

MACHINE LEARNING AND DATA PROCESSING CAPABILITIES

PRODUCTIVITY



ABLE TO BE MORE INNOVATIVE AND COMPETITIVE

RISK REDUCTION

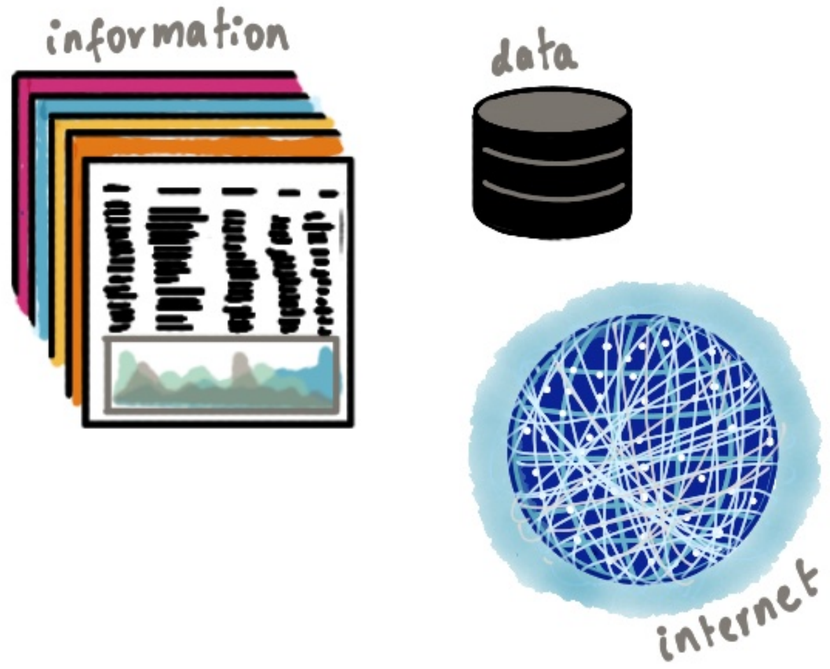
NO LONGER HAVING TO DO DULL, DIRTY OR DANGEROUS JOBS

WITH ALL THESE BENEFITS TO ENJOY, INTEREST AND INVESTMENT IN ROBOTS AND AUTONOMOUS SYSTEMS APPEARS TO BE TAKING OFF.

FEATURES

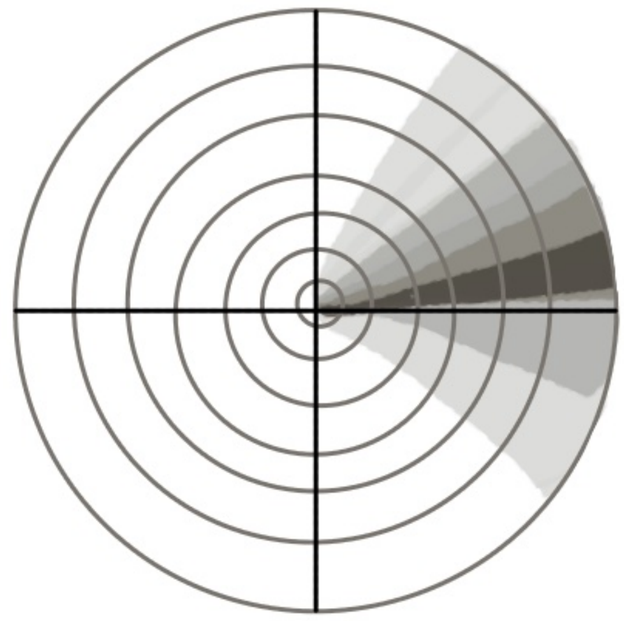
MAKE DECISIONS

USE DATA/INFORMATION

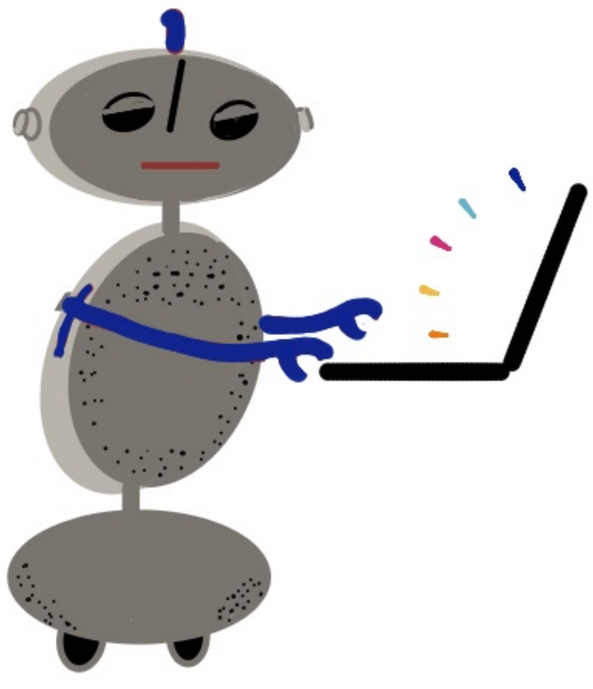


ARE LOCALISED OR DISTRIBUTED

INTERACT WITH SURROUNDINGS



ARE EMBODIED IN ROBOTS



AUTONOMOUS SYSTEMS

ACHIEVE A SET OF GOALS

IN A CHANGING ENVIRONMENT

AND WORK

FOR AN EXTENDED PERIOD

WITHOUT

HUMAN CONTROL OR INTERVENTION

EXAMPLES

DELIVERY DRONES

e.g. Wing, Prime Air, Zipline, ...

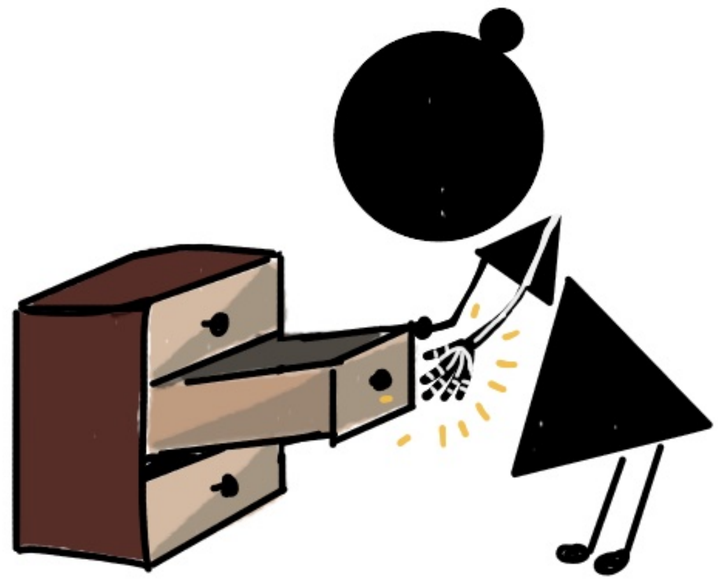
COMPANION ROBOTS

e.g. Pepper, Moxie, Lovot, Buddy, ...

UNDERWATER SURVEILLANCE

e.g. Raydrive in Royal Navy Uk

PROSTHETIC ARM

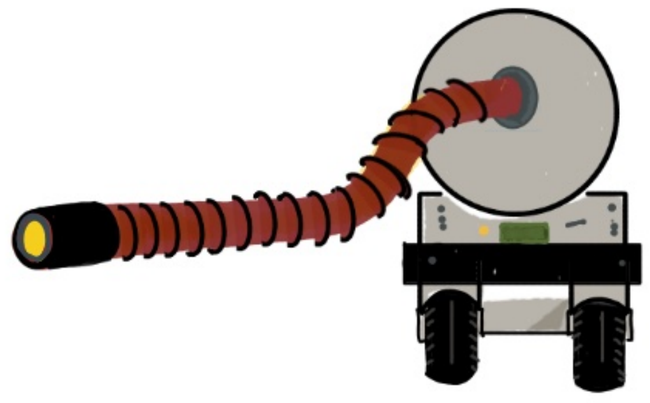


e.g Touch Bionics / Ossur

DRIVERLESS CARS

e.g. Tesla, Ford, Waymo, Pony.ai, ...

INSPECTION ARM



e.g. OC Robotics, Arm Hub, ...

ABILITIES

GATHER DATA

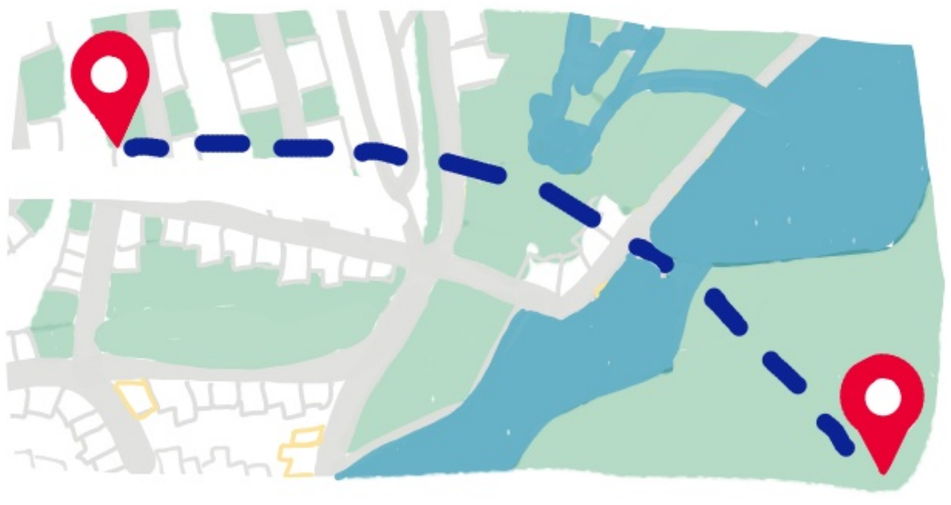
PROVIDE STATS ABOUT AN OBJECT, REGION OR PROCESS

EXAMPLE

MARINE CONSERVATION

TRANSPORT

CONVEY ITEMS OVER SHORT OR LONG DISTANCES.



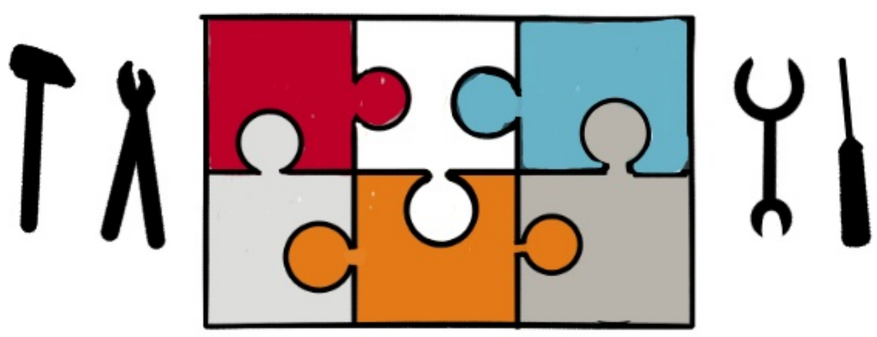
EXAMPLE

MATERIALS IN A FACTORY FLOOR

MEDICINES TO A REMOTE AREA

MANIPULATE

WORK WITH PHYSICAL OBJECTS AND TOOLS AS A HUMAN MIGHT



EXAMPLE

MANUFACTURING AND ASSEMBLY

SORT AND STORE

MANAGE INVENTORY - IDENTIFY, PACK, TRACK AND STORE ITEMS

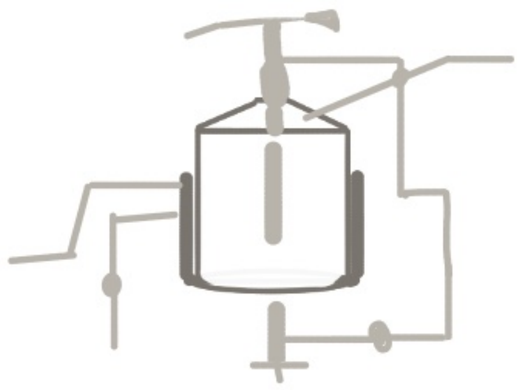
EXAMPLE

MEDICAL SAMPLES

RETAIL & MANUFACTURING

AUTONOMY FOR SAFETY

AUTONOMOUS SYSTEMS PROVIDE AN OPTION TO HELP WITH DECISION MAKING AND TO KEEP HUMANS OUT OF PHYSICAL HARM/INJURY. AN AUTONOMOUS SYSTEM COULD BE DEPLOYED IN PLACE OF A HUMAN IN THE FOLLOWING SCENARIOS



CHEMICAL REACTORS

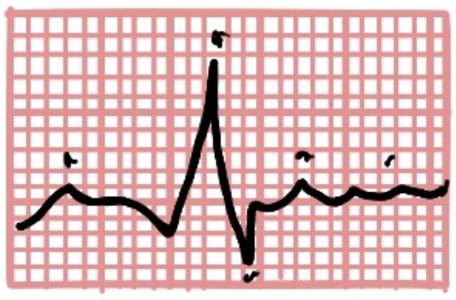
TO MONITOR CONDITIONS

EARTHQUAKE ZONES

TO LOOK FOR SURVIVORS

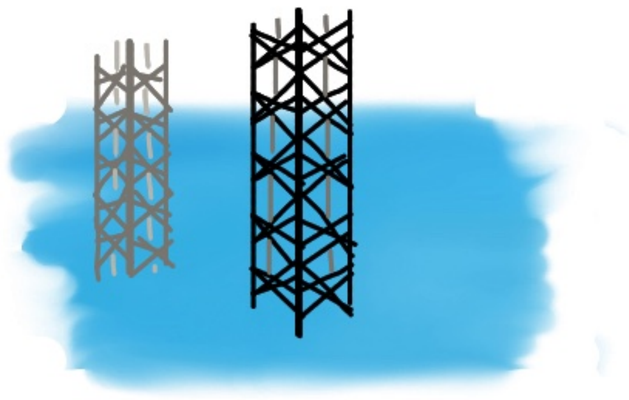
RESCUE OPERATIONS

TO DELIVER AID



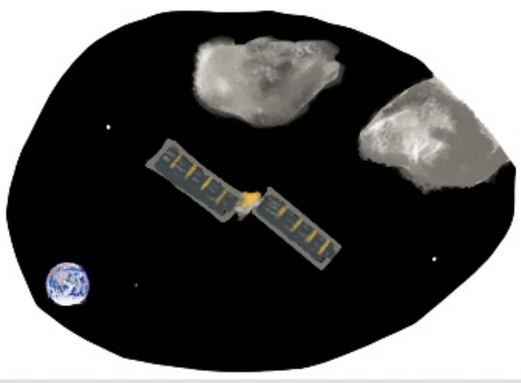
LIFE SAVING SURGERY

REMOTE SURGERY USING HAPTIC TECHNOLOGY



UNDER SEA

TO INSPECT ASSETS/MARINE LIFE



SPACE PROBES

IN EXPLORATIONS OR ASTEROID DEFLECTION

AUTONOMY IN WARFARE

MOST TECHNOLOGY IS DUAL-USE:

- USED FOR CIVILIAN/COMMERCIAL APPLICATIONS AND
- USE IN THE MILITARY.

UNSUPRISINGLY, AUTONOMOUS SYSTEMS ARE USED

- DURING PEACETIME TO PREPARE FOR CONFLICT AND
- DURING WAR

SOME EXAMPLES ARE:

TRAINING FOR BATTLEFIELD EXPERIENCE

WITH VIRTUAL/AUGMENTED REALITY

INTELLIGENCE ANALYSIS

WITH INFORMATION/LANGUAGE PROCESSING

CONTROLLING SEMI-AUTONOMOUS WEAPONS AND MISSILE GUIDANCE SYSTEMS

WITH COMPUTER VISION, GPS

PSYCHOLOGICAL WARFARE

BY EXPLOITING ECHO-CHAMBERS, DEEPFAKES

BEHIND THE SCENES

TECH INNOVATIONS HAVE THEIR RESPECTIVE STRENATHS AND LIMITATIONS. SOME CHALLENAES CAN BE MET WITH PROCESSES OR INVENTIONS. SOME OTHERS MIAHT REMAIN UNSOLVED.

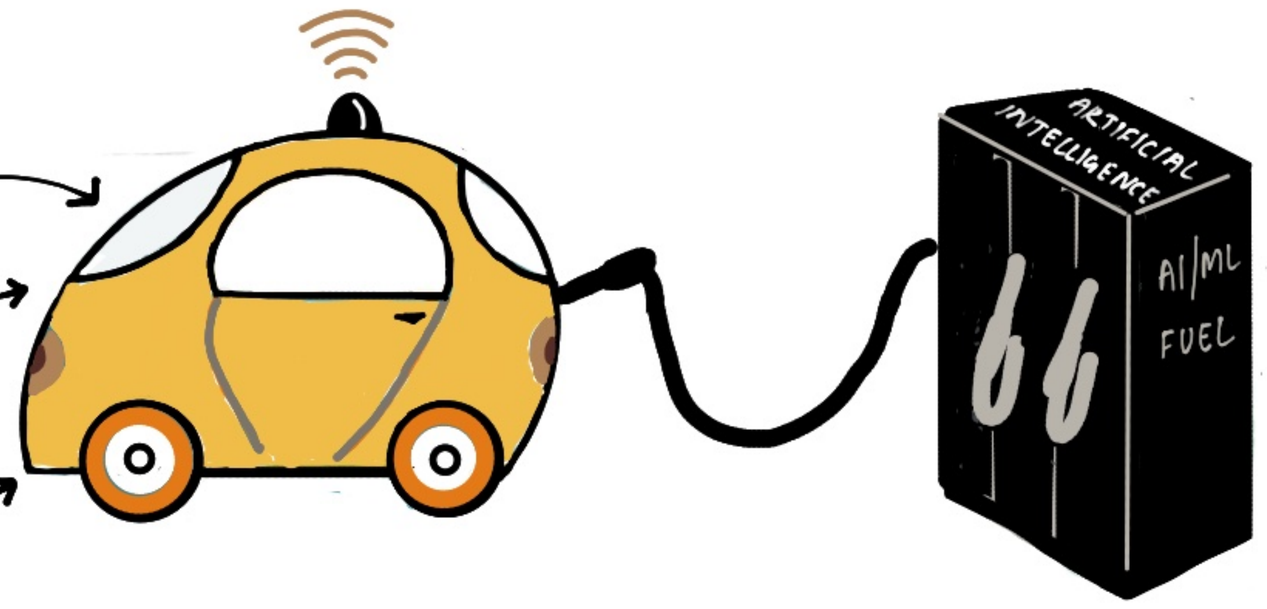
INTELLIGENCE

AUTONOMOUS SYSTEMS ARE POSSIBLE ONLY DUE TO THE IMPROVEMENTS IN ARTIFICIAL INTELLIGENCE.

AN INTELLIGENT AUTONOMOUS SYSTEM CAN ANALYSE A DEVELOPING SITUATION AND ACT TOWARDS ACHIEVING ITS GOALS SAFELY.

EXAMPLES

- NEAR VISION CAMERA
- RADARS & SENSORS
- GEO LOCATION
- ULTRASONIC SENSORS

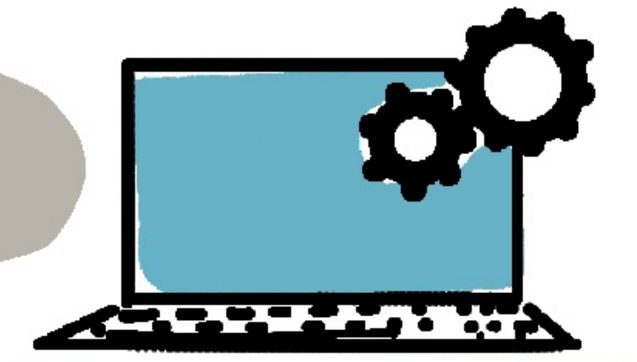


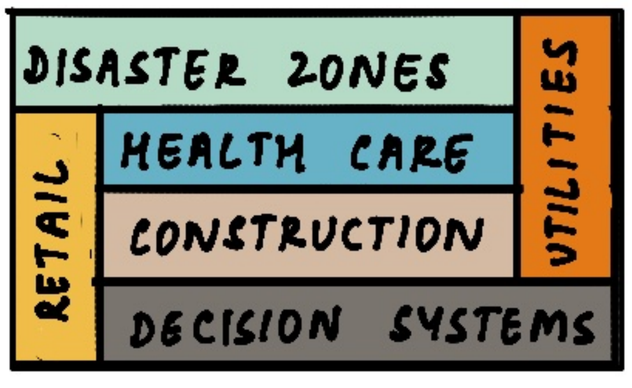
THE KIND OF METHODS REQUIRED

DEPENDS ON

THE PROBLEM BEING SOLVED

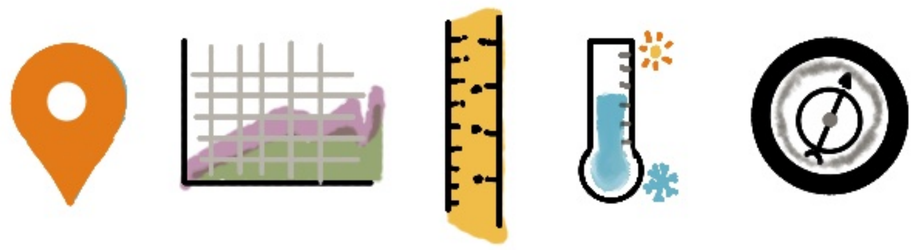
- SEMI SUPERVISED LEARNING
- REINFORCEMENT LEARNING
- APPROXIMATE DYNAMIC PROGRAMMING
- REGRESSION





DATA & ITS SEQUEL

AUTONOMOUS SYSTEMS COLLECT A HUGE AMOUNT OF DATA



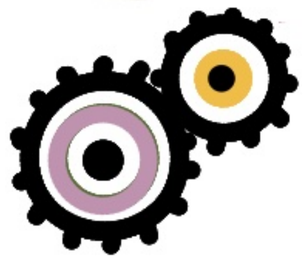
THIS DATA IS

STORED

FOR ANALYSIS/ FUTURE USE

SHARED

BETWEEN DEVICES



PROCESSED

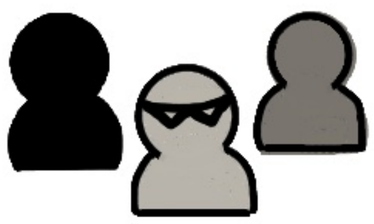
TO CONSOLIDATE UNDERSTANDING

INTERPRETED

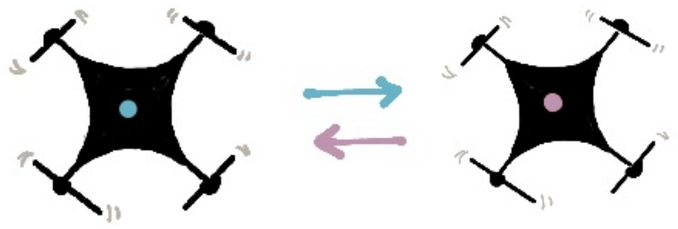
TO MAKE DECISIONS

CONSIDERATIONS

THE DATA COLLECTED LIVES OUTSIDE THE SOURCES AND LIKELY NOT WITH THE OWNER OF THE SOURCE ITSELF.



WHO OWNS AND ACCESSES THE DATA



HOW TO DEVELOP DATA SHARING STANDARDS

WHILE KEEPING PUBLIC SAFETY IN MIND

HOW TO PROTECT PRIVATE INFORMATION

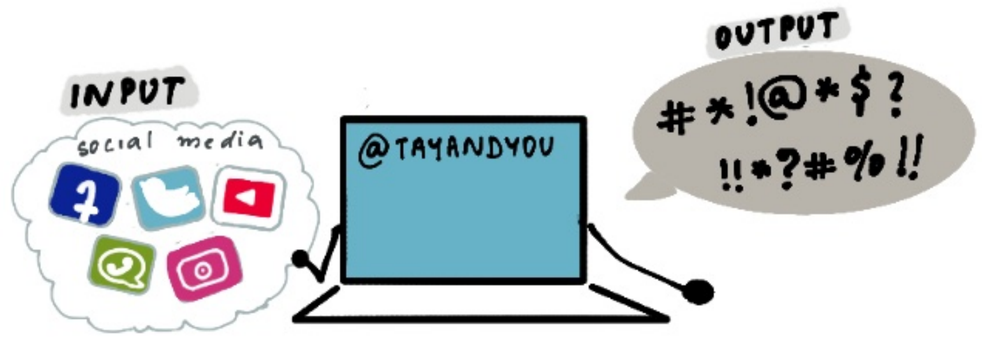
HOW TO TAP INTO THE VAST AMOUNT OF DATA

ENGINEERING CHALLENGES

UNCERTAIN ENVIRONMENTS

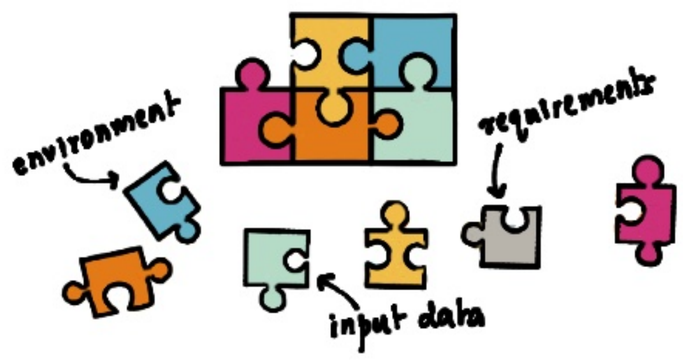
HOW TO ENSURE REASONABLE AND SAFE BEHAVIOUR IN NEW UNCERTAIN ENVIRONMENTS ?

SELF LEARNED NEW BEHAVIOUR



HOW TO ENSURE BEHAVIOUR IS CONSISTENT TO ORIGINAL SPECIFICATION AS NEW LEARNING OCCURS ?

INCOMPLETE SPECIFICATIONS

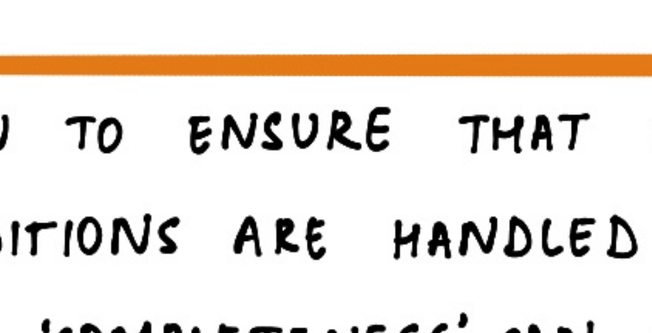


HOW TO ENSURE THAT RELEVANT CONDITIONS ARE HANDLED AND HOW 'COMPLETENESS' CAN IMPROVE ?

ETHICAL REASONING

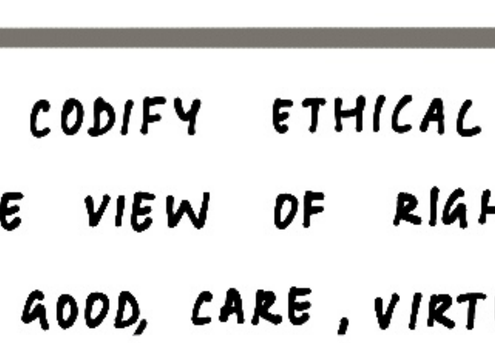
HOW TO CODIFY ETHICAL REASONING FROM THE VIEW OF RIGHTS, JUSTICE, COMMON GOOD, CARE, VIRTUE ETC ?

SCALING & ARCHITECTURE



HOW TO ENSURE ARCHITECTURES THAT CAN SCALE, USE RESOURCES AND VERIFY COMPLEX SYSTEMS ?

TRAINED & QUALIFIED PEOPLE



HOW TO ENSURE THAT THERE IS A GOOD SUPPLY OF QUALIFIED PROFESSIONALS FOR THE JOB ?

ON TESTING

AUTOMATED SYSTEMS

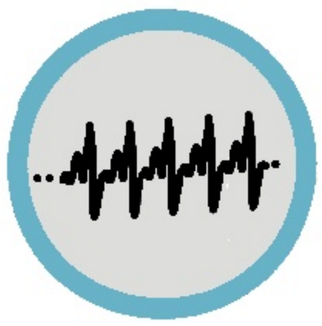
- FIXED SET OF RULES
- MOSTLY REPEATABLE
- FAIRLY DEFINED INPUT SPACE

AUTONOMOUS SYSTEMS

- NON-DETERMINISTIC
- COMPLEX / EMERGENT
- VAST INPUT SPACE

NEED DIFFERENT TESTING METHODS

AUTONOMOUS SYSTEMS



MAKE CRITICAL DECISIONS

OPERATE IN HIGH-RISK SITUATIONS

THEREFORE, QUALITY TESTS, SAFETY TESTS } ARE IMPORTANT, CHALLENGING

BLACK BOX PROBLEM

THE AI/ML CODE POWERING AUTONOMOUS DEVICES TEND TO BE VERY HARD TO EXPLAIN. SO, VERIFICATION & VALIDATION PROCESSES NEED TO

- MAKE THE ALGORITHM TRANSPARENT
- TRACE THE DECISION MAKING PROCESS
- PROVE RELIABILITY WHEN UPGRADED

THIS WILL GO PART WAY TO ANSWERING SOME OF THE CONCERNS UNDER PERFORMANCE AND PROCESS IN PROVING TRUST WORTHINESS

AUTONOMY TESTING TASKS

SAFETY AND RELIABILITY

KNOW THAT THE GOAL(S) ARE MET AND IT IS SAFE FOR ALL

EVIDENCE FOR CERTIFICATION

UNDERSTAND WHAT DECISION WAS MADE AND WHY

ETHICAL BIAS

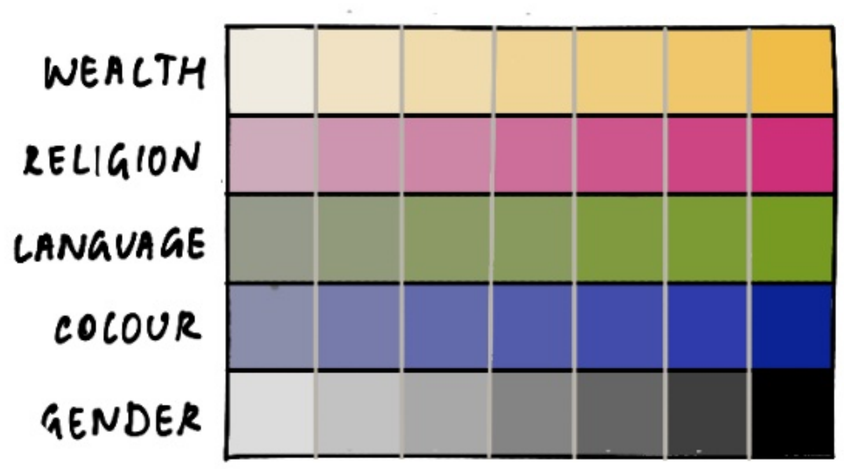
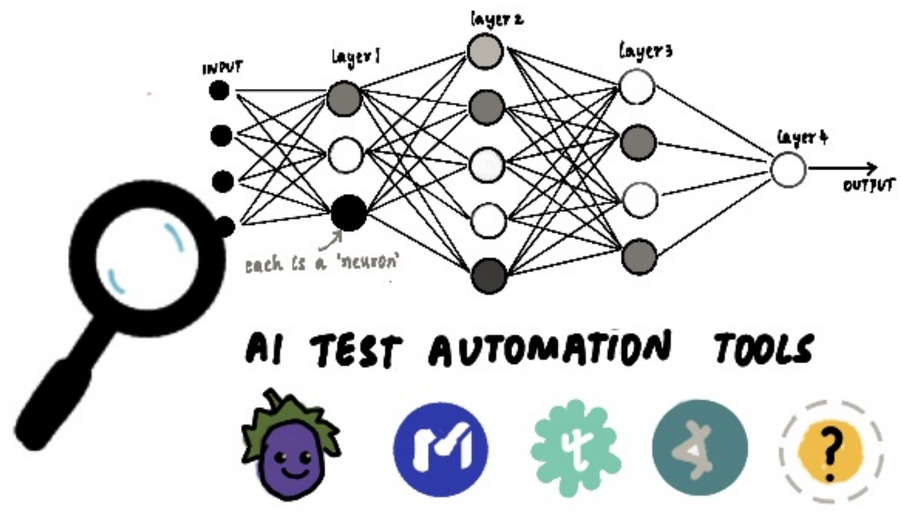


FIGURE OUT WHAT SOCIAL OR CULTURAL BIASES ARE EXHIBITED

TOOLS AND TECHNOLOGY



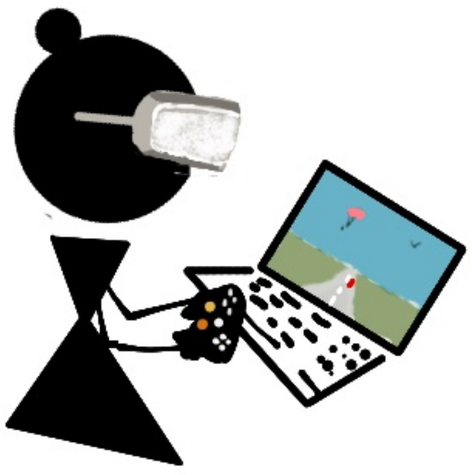
DEVELOP FIT-FOR-PURPOSE TEST TOOLS AND CONTINUOUSLY EVALUATE

TESTING METHODS

THERE ARE A FEW METHODS TO TEST AUTONOMOUS SYSTEMS.

WE WILL LOOK VERY BRIEFLY AT JUST THREE OF THOSE HERE.

SIMULATIONS . . .



- STAND IN PLACE OF THE REAL / PHYSICAL WORLD
- USE BOTH HARDWARE AND SOFTWARE

COGNITIVE TESTS . . .

..RUN A SET OF WELL-DEFINED

List of :
- Scenarios
- Disruptions
- Constraints

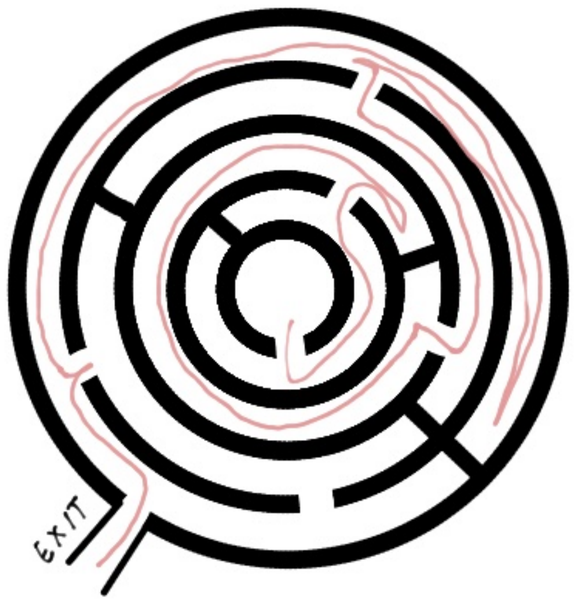
AND COMPARE THE OUTCOMES TO A SET OF

Expectations:
- Law
- Experience
- Expertise
- Ethics / norms

TO MAKE THE IMPLICIT RULES IN THE ALGORITHM TRANSPARENT

SHORT FIELD TESTS

INVOLVE



USING THE SYSTEM IN A CONTROLLED REAL SETTING

RECORDING DATA FROM ALL SENSORS

LOGGING ALL THE OPTIONS IT CONSIDERED

TO MAKE THE FEEDBACK / REWARD MECHANISMS TRANSPARENT & OPTIMISE OFFLINE

CHALLENGES IN TESTING

TRACKING DECISIONS

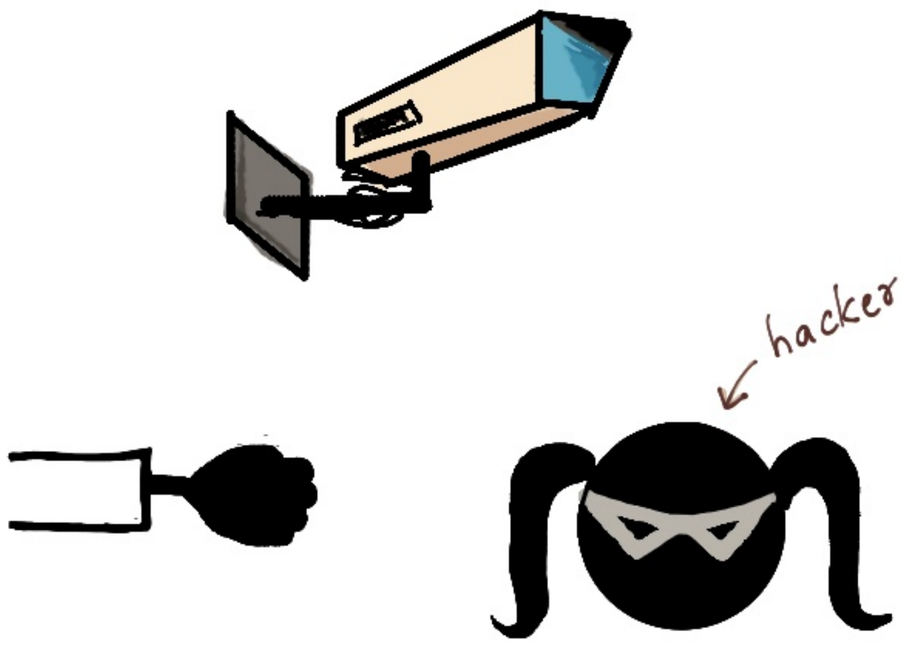
HOW TO TRACK AND UPDATE DECISIONS AS THE SYSTEM LEARNS OR GETS UPGRADED?

UNCLEAR REQUIREMENTS

- IMPLICIT / UNSTATED
- OPEN TO INTERPRETATION
- CONTRADICTING GOALS
- TENDING TOWARDS A BIAS

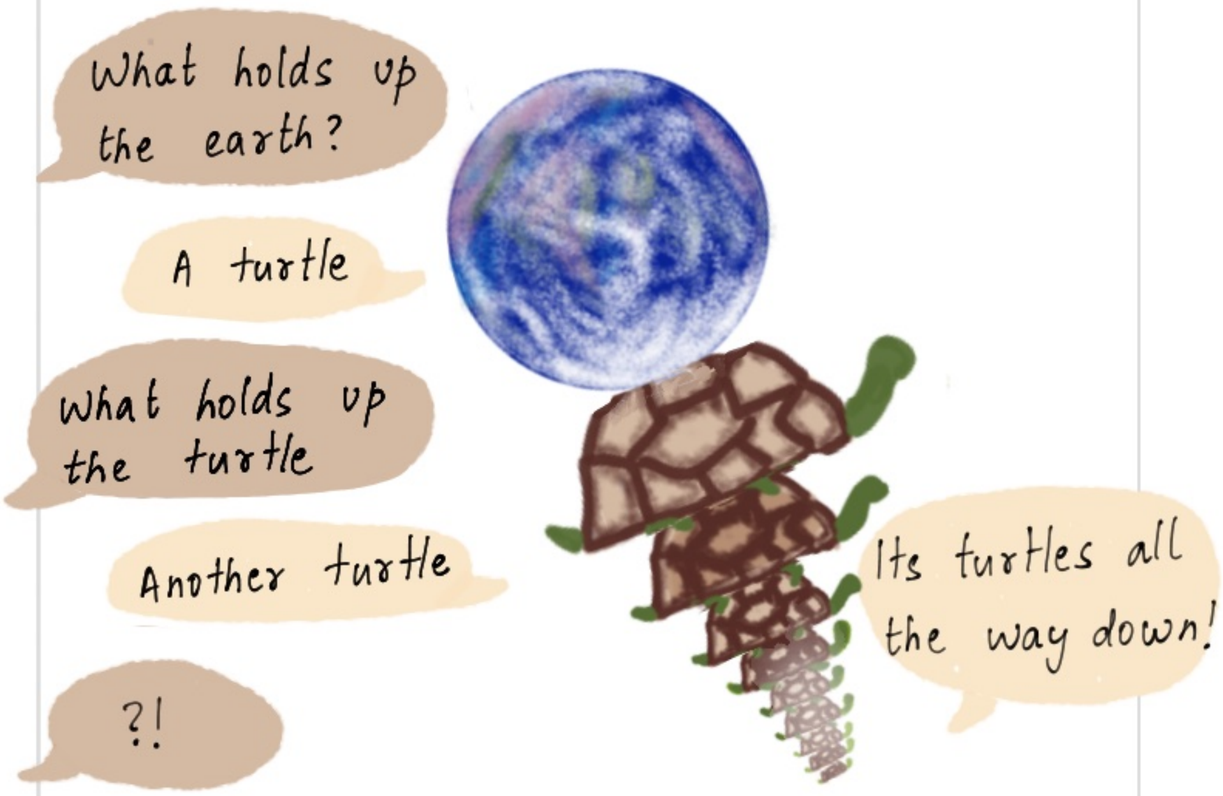
HOW TO PROVE CORRECTNESS GIVEN ANY OF THESE IS TRUE?

SECURITY CHALLENGES



IN THE ABSENCE OF CONTINUOUS SUPERVISION, HOW TO TEST FOR SELF-DEFENCE AGAINST ATTACKS?

TESTING VERIFICATION TOOLS



HOW TO PROVE CORRECTNESS OF THE VERIFICATION TOOLS?

ON STANDARDS REGULATION AND LAW

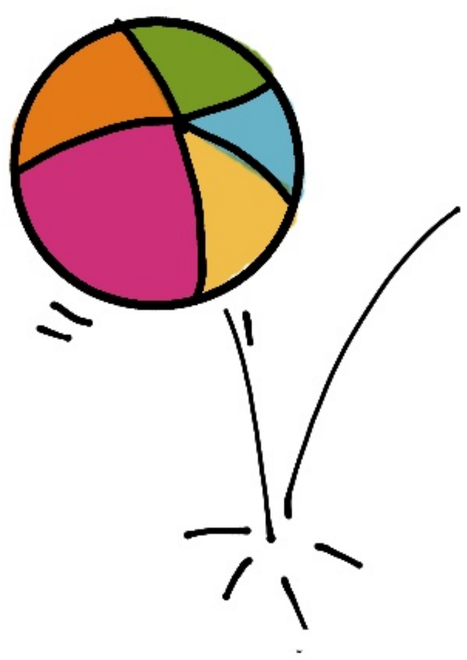
REGULATIONS RELY ON SAFETY AND RELIABILITY STANDARDS. REGULATED PRODUCTS PERFORM WELL-DEFINED OPERATIONS IN WELL-UNDERSTOOD CIRCUMSTANCES. AUTONOMOUS SYSTEMS UPEND ALL THESE NOTIONS.

ON STANDARDS

HOW A STANDARD IS USUALLY DEFINED

A PERFORMANCE STANDARD HAS THREE ASPECTS

FUNCTION

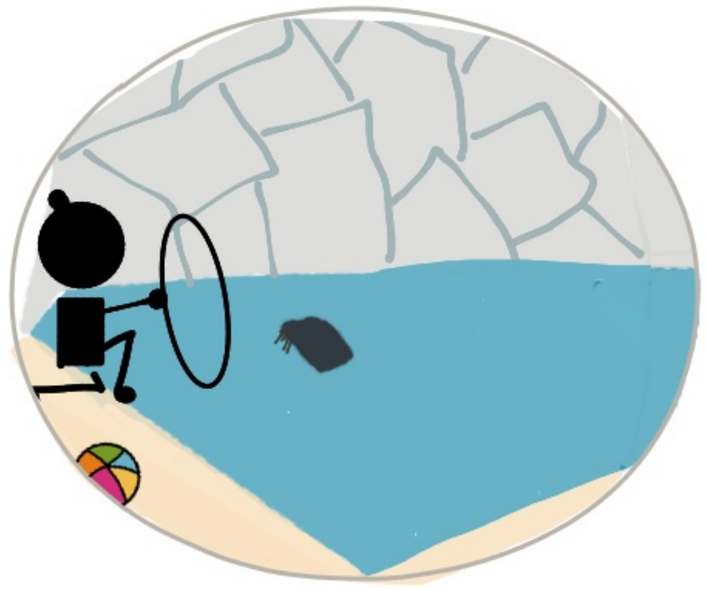


WHAT THE TECH DOES TO ACHIEVE ITS GOAL

TOLERANCE

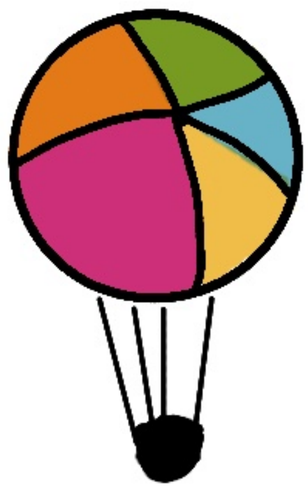
HOW OFTEN/CLOSELY IT MUST ACHIEVE ITS FUNCTION

CONTEXT



ENVIRONMENTS IN WHICH THE TECH IS MEANT TO OPERATE

WHY THIS IS A CHALLENGE WITH AUTONOMOUS SYSTEM



AUTONOMOUS SYSTEMS STUDY THE CONTEXT AND ADAPT THEMSELVES TO DISPLAY BEHAVIOURS SUITABLE TO THE ENVIRONMENT AND ARE NOT LIMITED BY FUNCTION, TOLERANCE OR CONTEXT. THIS LACK OF BOUNDARY BECOMES PARTICULARLY CONCERNING IN LIFE-AND-DEATH SITUATIONS.

FOR EXAMPLE:

- SELF-DRIVING CARS ON ROADS WITH PEDESTRIANS/WILDLIFE
- AUTONOMOUS WEAPONS/MISSILES IN A CIVILIAN AREA

ON REGULATION

REGULATION

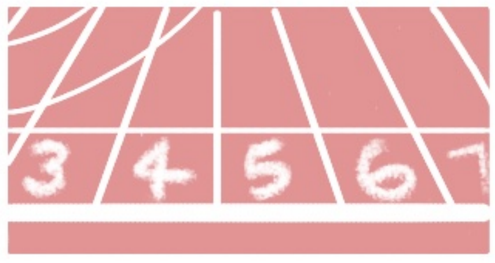
A RULE OR MECHANISM THAT LIMITS, STEERS OR OTHERWISE CONTROLS CERTAIN BEHAVIOURS.

ENFORCEABLE BY LAW

LIMITED TO THE SCOPE OF THOSE UNDER THE AUTHORITY THAT REGULATES THEM.

REASONS TO REGULATE

CUSTOMERS

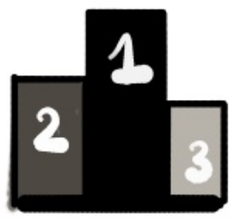


COMPETITION

- TO PROTECT THE INTERESTS OF CUSTOMERS OF THE SERVICE AND OF COMPETITORS AND
- TO ESTABLISH ACCOUNTABILITY

CONSIDERATIONS

OBJECTIVES OF THE SYSTEM



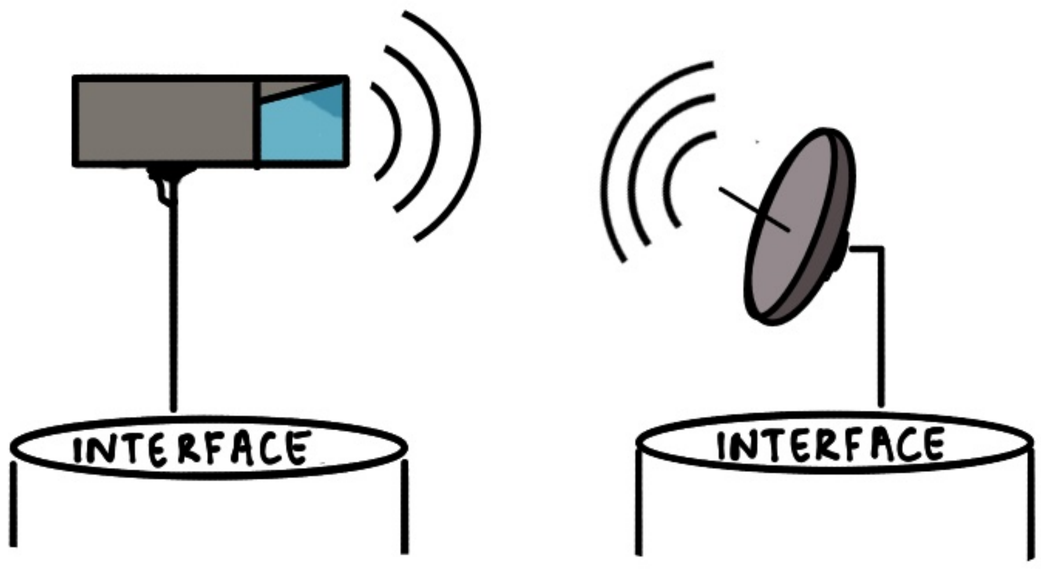
STANDARDS TO EVALUATE PERFORMANCE

POTENTIAL RISK POSED BY THE COMPANIES

COLLABORATION WITH THE PRIVATE SECTOR

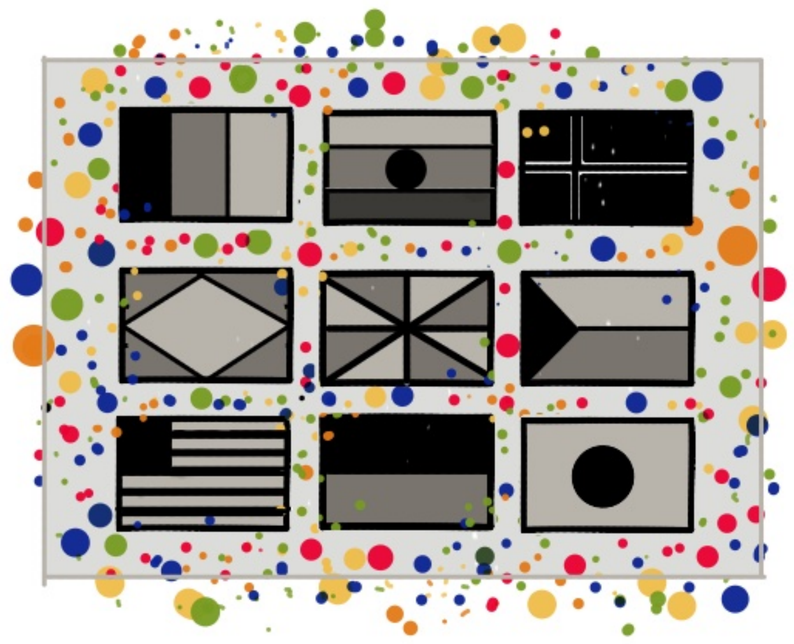
REGULATORY CHALLENGES

PROPRIETARY STRUCTURES



HOW TO REGULATE INTERFACES TO AUTONOMOUS SYSTEMS LEAVING LOW-LEVEL STRUCTURES TO STAY PROPRIETARY ?

DIFFERENT COUNTRIES

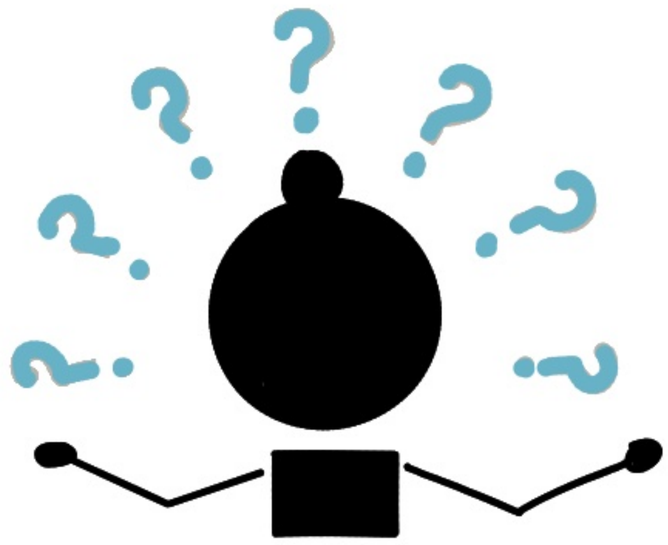


HOW TO ENSURE AUTONOMOUS SYSTEMS COMPLY WITH DIFFERING OR CONTRADICTING LAWS IN EVERY COUNTRY IT OPERATES ?

CONFLICTING POSITIONS

HOW CAN REGULATION STRIKE A BALANCE BETWEEN HELPING INNOVATION AND CURBING MALICIOUS TECHNOLOGY ?

WHAT WE DON'T UNDERSTAND



HOW TO REGULATE WHAT WE DO NOT UNDERSTAND - A TECHNOLOGY THAT IS RELATIVELY NEW ?

HOW TO REGULATE

A PAPER PUBLISHED IN IEEE

Regulating Autonomous Systems: Beyond Standards

DAVID DANKS

ALEX JOHN LONDON

Carnegie Mellon University

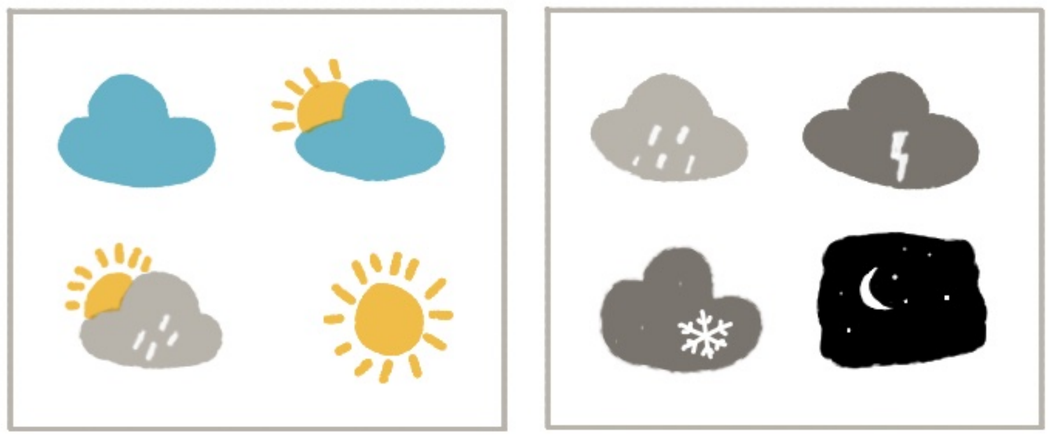
PROPOSES TWO OPTIONS

- LIMITING THE SCOPE OF AUTONOMY
- USING A MORE DYNAMIC METHOD AS WITH PHARMACEUTICALS

LIMIT THE SCOPE

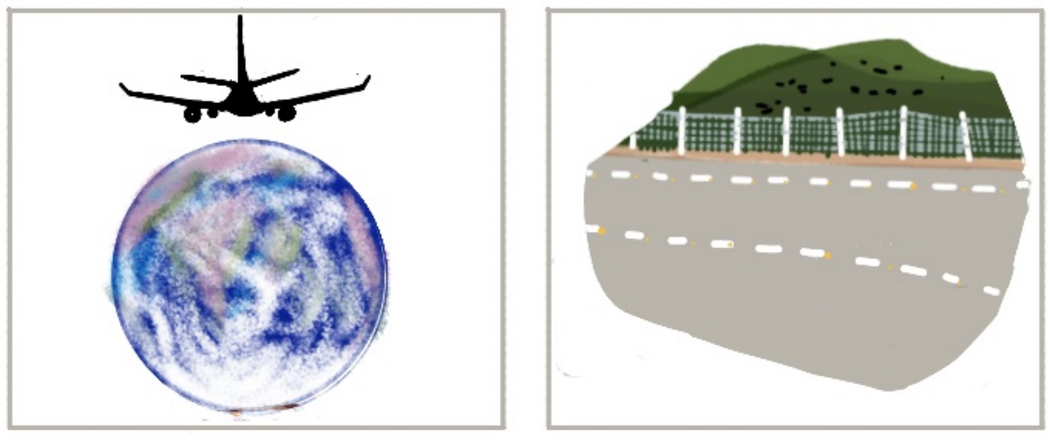
LIMITING THE SCOPE OF AUTONOMY ENABLES THE USE OF MORE CONVENTIONAL APPROACHES TO REGULATION

BY MAKING HUMANS IDENTIFY CONTEXT



NOT IDEAL, BUT USED IN SEMI AUTONOMOUS DRIVING

BY REGULARISING THE ENVIRONMENT



METHODS SUCH AS:

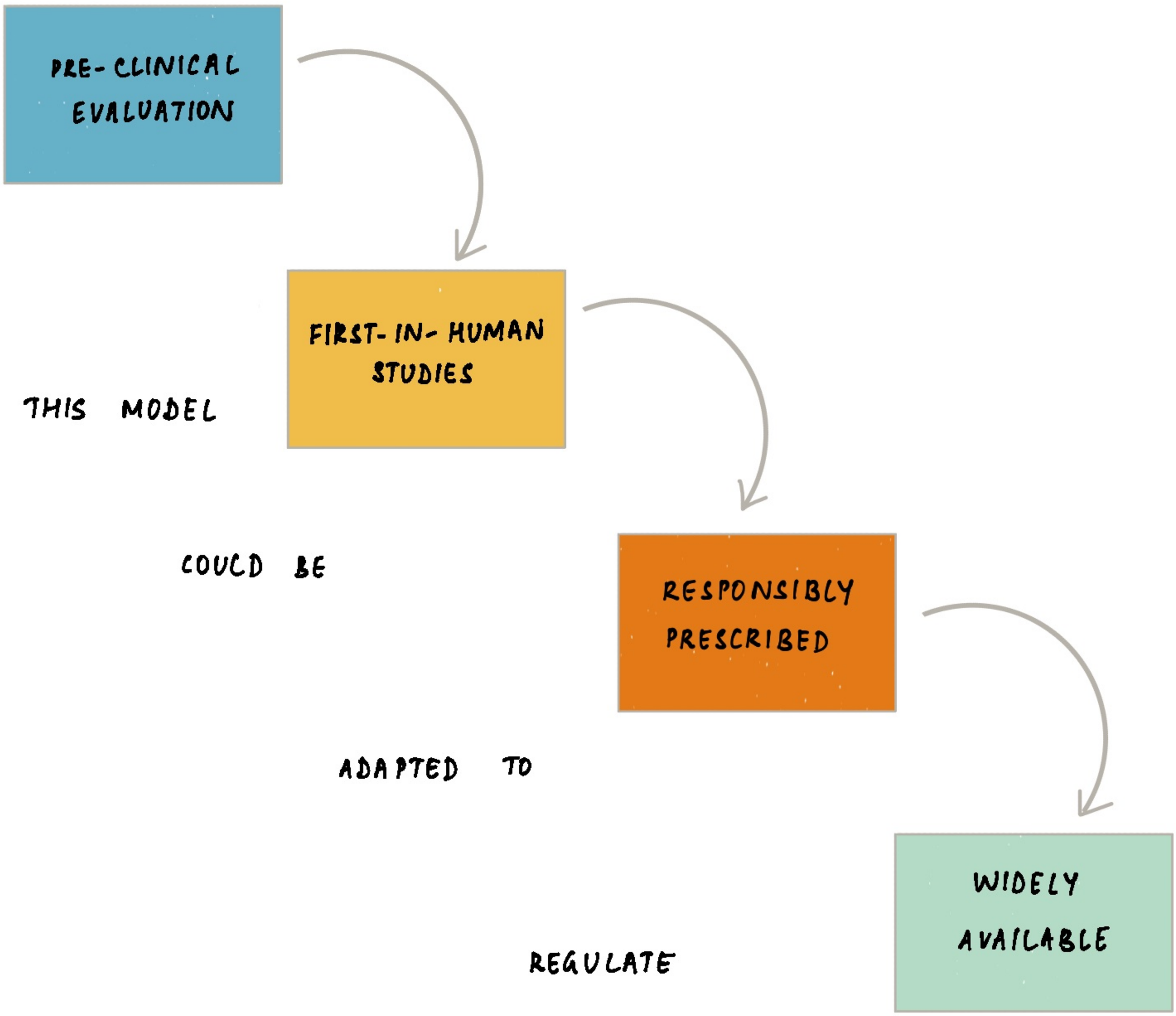
- LIMITED ACCESS POINTS FOR AIR TRAFFIC
- ANIMAL FENCING ON HIGHWAYS

REDUCE NEED TO CLARIFY CONTEXT

UNFORTUNATELY, THESE TWO APPROACHES STOP US ENJOYING THE PROMISED BENEFITS OF AUTONOMY.

A MORE DYNAMIC METHOD -1

THE PHARMACEUTICAL DRUG APPROVAL PROCESS IN STAGES IS ROUGHLY



THIS MODEL COULD BE ADAPTED TO REGULATE AUTONOMOUS SYSTEMS.

IT REQUIRES

- CONTINUOUSCY GATHERING DATA
- A REGULATORY BODY THAT EVALUATES/BENCHMARKS

A MORE DYNAMIC METHOD - 2

WHEN APPLIED TO AUTONOMOUS SYSTEMS, THE STAGES MIGHT BE:

STAGE	GOAL: TO FIND
TESTING IN SIMULATED ENVIRONMENTS	NEW VARYING SITUATIONS • HOW IS DATA USED? • HOW ARE DECISIONS MADE? • HOW ARE RULES/CONSTRAINTS OBEYED?
LIMITED TARGETED REAL-WORLD SETTING	 MONITORED BY TRAINED USERS • BEHAVIOUR IN REAL SETTING • CAN CONTEXT RECOGNITION IN HARDWARE/SOFTWARE IMPROVE?
REPEAT IN A DIFFERENT SETTING	 • WHAT MAKES IT NEW? • HOW TO MONITOR IT?
RELAXING RESTRICTIONS ON MARKET ACCESS	 SUCCESS RELIES ON DETECTION & RESPONSE • SOLUTIONS, IMPROVEMENTS • DEGREES OF AUTONOMY • ANY LONG TERM RELIABILITY PATTERNS

ON LAW

WHY CARE ABOUT LAW ?

AUTONOMOUS SYSTEMS IMPACT

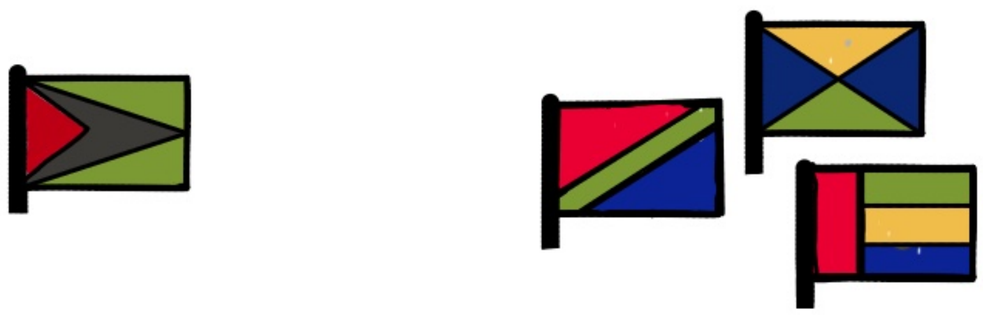
- RISKS & OPPORTUNITIES
- RIGHTS & DUTIES
- SAFETY & ACCOUNTABILITY

OF INDIVIDUALS & ORGANISATIONS

CONSIDERATIONS

- IS ITS USE ETHICAL ?
- IS ITS USE MORAL ?
- WILL EXISTING LAWS SUFFICE ?
- IF NOT, WHAT TO DO?

COMMERCIAL USE



- MOSTLY DOMESTIC LAWS APPLY FOR CIVILIAN USE PRODUCTS
- INTERNATIONAL LAW RELEVANT ONLY FOR CONSISTENCY

MILITARY USE

INTERNATIONAL HUMANITARIAN LAW

- IMPOSES LIMITS ON SUFFERING CAUSED BY ARMED CONFLICT
- NO BAN YET ON AUTONOMOUS WEAPONS OR KILLER ROBOTS

SELF-REGULATION ?

↓

WILL NOT SUFFICE - GIVEN PRIVATE COMPANIES' HANDLING OF

- FAKE NEWS
- MANIPULATIVE ALGORITHMS
- BIASED/ABUSIVE CONTENT

IN BALANCE ...

IF THE LAW IS TOO SPECIFIC

→ GETS OUTDATED QUICKLY

IF THE LAW IS TOO GENERAL

→ INSUFFICIENT GUIDANCE

WHO DO WE TRUST?

LIFE ALTERING DECISIONS ARE TO BE MADE BY INDEPENDENTLY OPERATING MACHINES. WHO DECIDES THAT THESE DECISIONS ARE 'RIGHT' OR 'MORAL'? HOW CAN THE PUBLIC TRUST MACHINES?

ON TRUST - 1

WHAT IS TRUST ?

- A BELIEF
- AN EXPECTATION
- AN EMOTIONAL STATE OF THE MIND
- AN INTERPERSONAL FEELING

TRUST IS USUALLY BETWEEN PEOPLE. WHAT HAPPENS WHEN WE NEED TO TRUST A MACHINE? WHAT RELATIONSHIP WORKS?

BASED ON WHAT WE HAVE SEEN OF AUTOMATION OVER THE YEARS AND MORE RECENTLY, WITH ARTIFICIAL INTELLIGENCE, WE KNOW TRUST IS NOT EASILY EARNED.

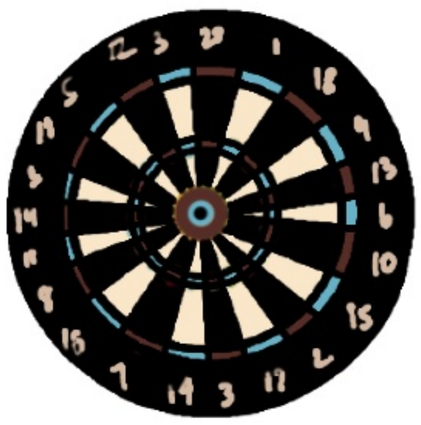
ON TRUST - 2

HOW THEN CAN TECHNOLOGY EARN THE TRUST OF THE PUBLIC? ACCORDING TO JACK STILGOE (UCL) & GOPAL RAMCHURN (UKRI), ONE WAY IS TO THINK ABOUT TECHNOLOGY IN THESE TERMS:

PEOPLE

WHO **OWNS** IT AND WHO **BENEFITS**?
WHOSE **INTERESTS** ARE PROTECTED?
WHO IS **LIABLE** WHEN THINGS GO WRONG?

PERFORMANCE



WHAT **VALUE** DOES IT PROVIDE?
WHAT **HUMAN VALUES** DRIVE IT?
WHAT ARE ITS **BOUNDARIES** AND **RISKS**?

PURPOSE

WHY DOES IT **EXIST**?
WHAT **PURPOSE** DOES IT SERVE?

PROCESS

WHAT PROCESS **SAFEGUARDS** | **REGULATES** IT?
WHICH **INSTITUTIONS** ARE INVOLVED?
HOW ARE THESE **COMMUNICATED**?

WHO DO WE BLAME?

MADELEINE CLARE ELISH, RESEARCH SCIENTIST, DESCRIBES THE DISPARITY BETWEEN THOSE WITH CONTROL AND THOSE WITH RESPONSIBILITY.

THE TECH & PROCESSES SEEM TO GET AWAY...

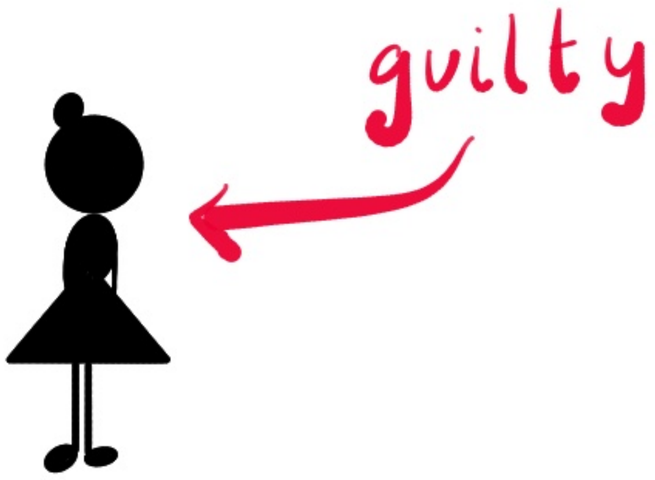
HUMAN IN THE LOOP

SELF-DRIVING CAR KILLS PEDESTRIAN

ARIZONA

MAR 18, 2018

DURING A TEST-DRIVE, AN UBER SELF-DRIVING CAR COLLIDED AND KILLED A CYCLIST ON FOOT. THE CAR WAS UNABLE TO CORRECTLY IDENTIFY THE WALKING CYCLIST AS A PERSON.



THE BACK UP SAFETY DRIVER WAS CHARGED WITH MANSLAUGHTER

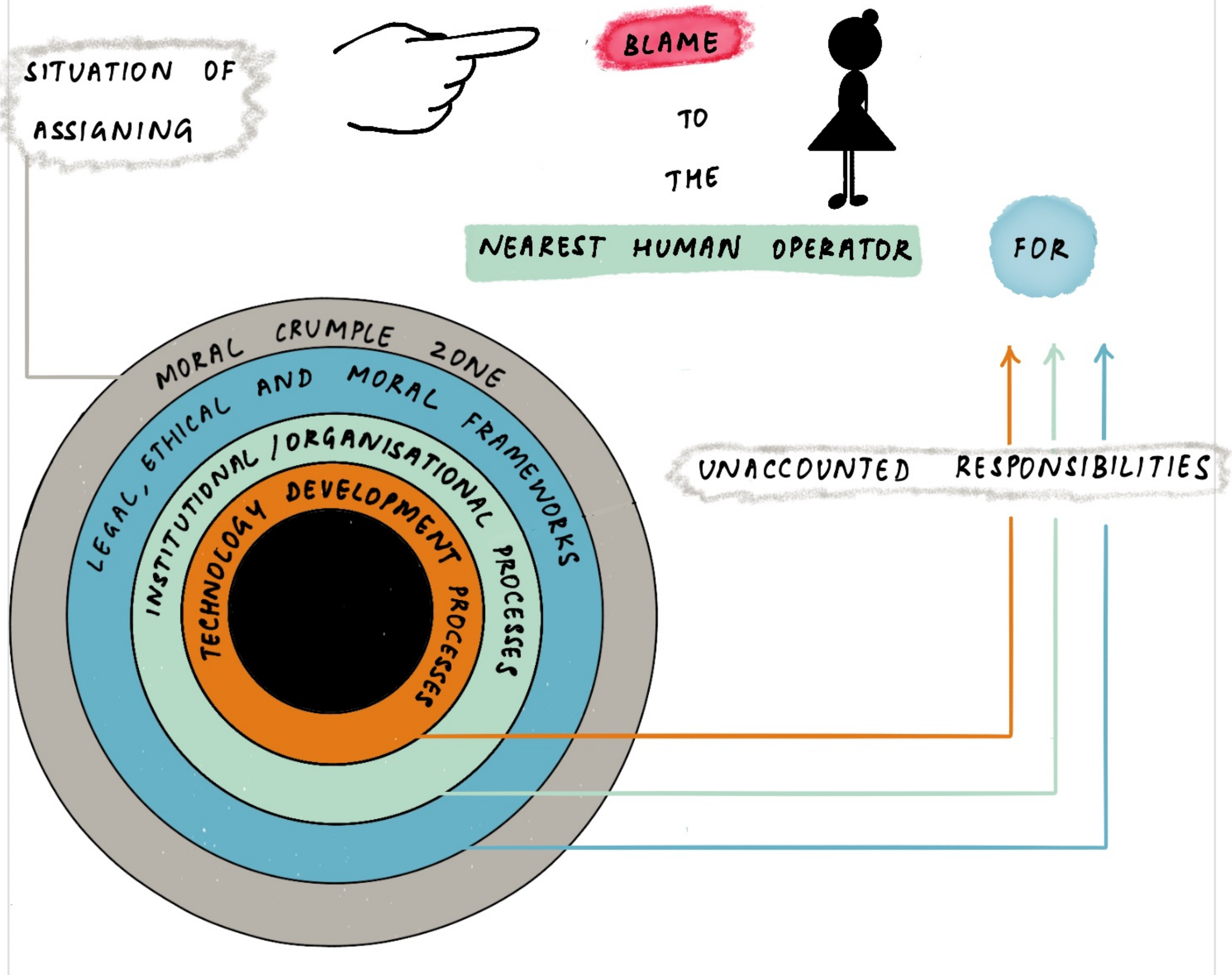
UBER WAS CLEARED OF WRONG-DOING.

HAVING A 'HUMAN IN THE LOOP' WAS A WAY TO ENSURE SAFETY, BUT THE EASIEST ONE TO BLAME.

MORAL CRUMPLE ZONE

WHO IS ACCOUNTABLE WHEN THINGS GO WRONG WITH TECHNOLOGY? CORPORATE AND LEGAL LIABILITY/RESPONSIBILITY HAVE NOT YET CAUGHT UP TO THE ADVANCES IN TECHNOLOGY.

MORAL CRUMPLE ZONE IS A PHRASE USED IN THIS CONTEXT.



UNFORTUNATELY, SOMETIMES IT HAPPENS THAT TECHNOLOGY IS VIEWED AS FAULTLESS TO THE DETRIMENT OF THE ESSENTIAL WORKERS OF TECHNOLOGY WHO HELP INTEGRATE IT INTO OUR LIVES.

THIS REMAINS AN OPEN PROBLEM WITH AUTONOMOUS SYSTEMS TOO.

DO WE CARE?

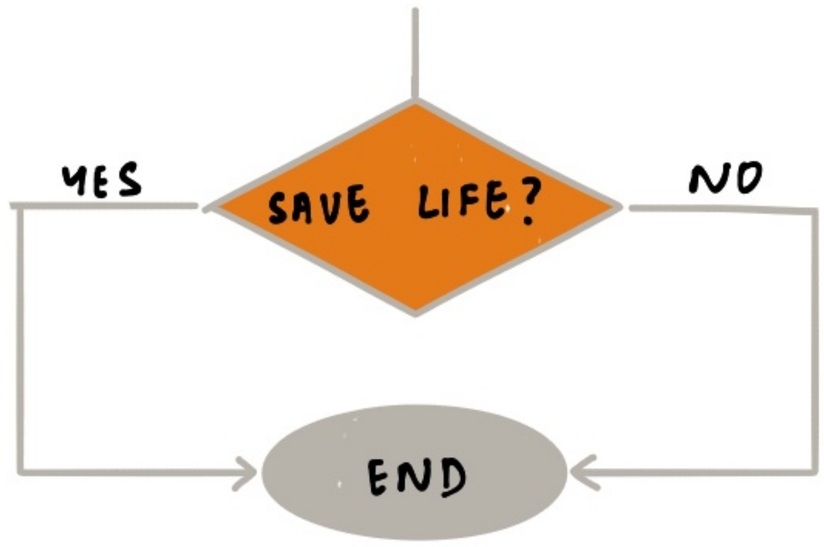
CYBERPHYSICAL DEVICES (DEVICES WITH INTELLIGENCE) ARE LIKELY TO BE EVERYWHERE. HOW DO WE PREPARE OURSELVES TO ACCOMODATE THEM IN DAILY LIFE? HOW DO WE FIND OUT?

ON ETHICS / DANGERS

JOBS

AUTOMATION CREATES JOBS THAT PREVIOUSLY DIDN'T EXIST. BUT HOW TO JUSTIFY BUILDING TECHNOLOGY THAT MAKES PEOPLE LOSE THEIR LIVELIHOOD?

LIVES



WE ARE MORE TOLERANT OF A HUMAN'S ERROR THAN OF MACHINE'S

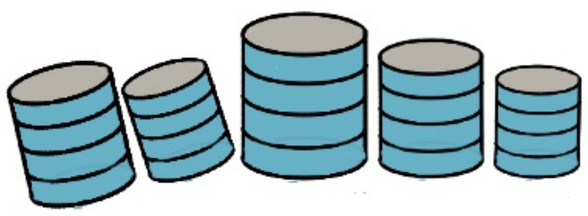
HOW TO RECONCILE WITH AN ALGORITHM THAT DECIDES WHO SURVIVES AND WHO DOES NOT?

DATA

"DATENSPARSAMKEIT"

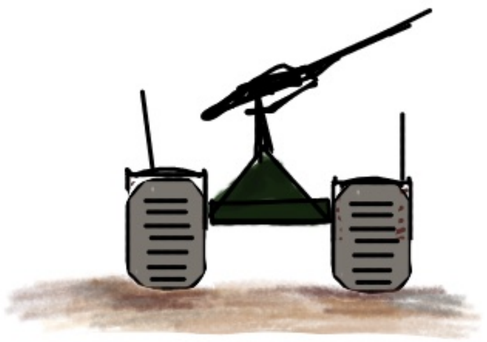
TOO MUCH DATA?

IS THERE CONSENT?



DEVICES POSSESS SO MUCH OF PERSONAL DATA, LEAVING US WITH SECURITY RISKS (OF WHICH WE MIGHT NOT EVEN BE AWARE)

RESPONSIBILITY



WHEN STATES USE AUTONOMOUS WEAPONS AGAINST CIVILIANS, WHO CHARGES THEM WITH CRIMINAL INTENT?

LIVING WITH AI

People + AI (EVERYWHERE) coexisting...

... with social and cultural implications...

GENEVIEVE BELL

PROF BELL POSES 5 QUESTIONS TO ASK, WHOSE ANSWERS MAY PREPARE US BETTER FOR THE FUTURE

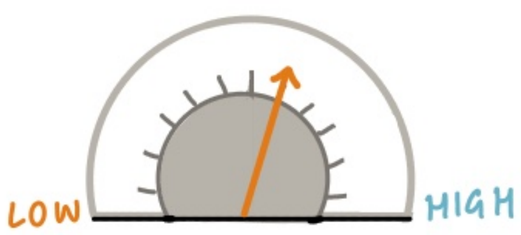
IS IT AUTONOMOUS?

IS A DEVICE SMART?

IS IT AUTONOMOUS?

HOW CAN WE TELL?

DOES IT HAVE AGENCY?



ARE THERE LIMITS/CONTROLS FOR ITS FUNCTIONS?

WHO DECIDES?

WHO EXERCISES?

CAN WE HAVE THE ASSURANCE?

TRUST

PRIVACY

LIABILITY

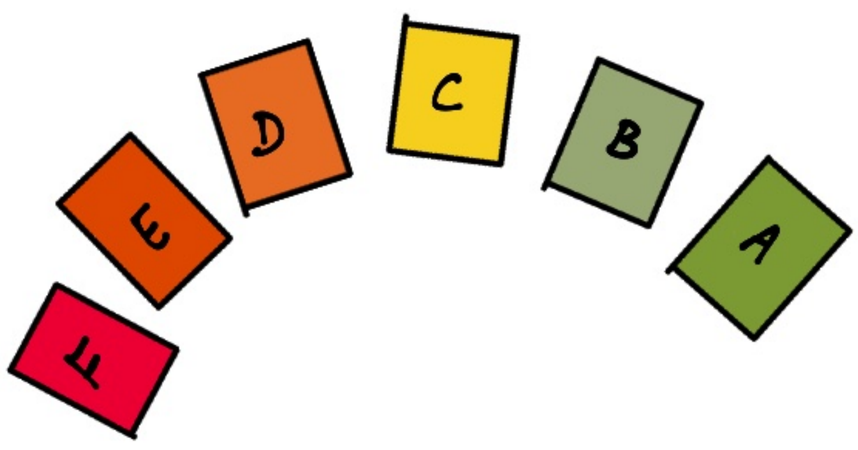
ETHICS

WHAT IS TOLERABLE?

WHOSE SAFETY?

WHO DECIDES?

HOW TO MEASURE SUCCESS?



IS IT LIVES SAVED/LOST?

WHAT ARE OTHER INDICATORS?

CAN IT INTERACT HUMAN-LIKE?

FINGERPRINT

VOICE

****** PASSWORD

WHAT MEANS OF INTERACTION FIT?

FOR ONE? FOR MANY?

A BRIEF OUTLOOK

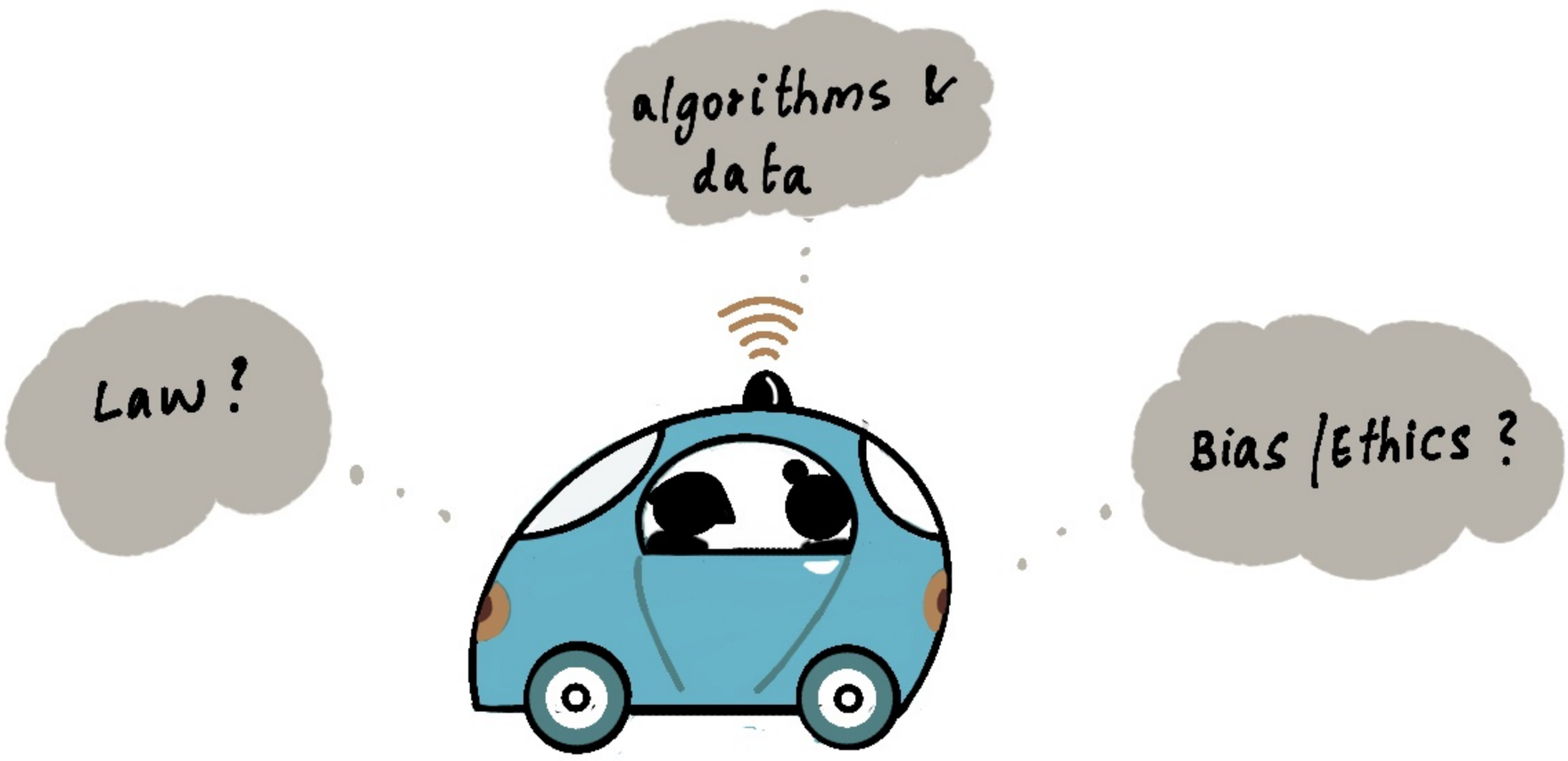
PROF DAVID LANE IMAGINES A FUTURE WHERE AUTONOMOUS SYSTEMS CAN BE THOUGHT OF LIKE A PHONE WITH MANY APPS: ONE 'ROBOT' WITH MANY SKILLS.

AUTONOMY OF MACHINES IS HERE TO STAY.

HOWEVER, AUTONOMOUS SYSTEMS HAVE MANY LIMITATIONS.

THEY ARE CONSTRAINED BY THE TECHNOLOGY AVAILABLE TO BUILD OR BETTER THEM. THEY HAVE LIMITED FUNCTIONALITY.

THEY FAIL AT THE HANDOVER POINTS BETWEEN MACHINE AND HUMAN. THEY HAVE NO MORALS OR ETHICS OF THEIR OWN.



THEIR DECISION MAKING IS BIASED BY THE HUMANS THAT CODED/TRAINED THEM.

PERHAPS, THEY CAN NEVER BE TRULY AUTONOMOUS, AS THEY WILL NEVER OPERATE INDEPENDENTLY OF HUMANS.

MORE TO EXPLORE

THERE ARE MANY TOPICS THAT THIS GUIDE DOES NOT COVER. HERE IS A SMALL LIST OF INTERESTING RESEARCH TO FURTHER PIQUE INTEREST.

- MACHINE TEACHING
  HUMAN EXPERTISE GIVEN TO MACHINES

- REINFORCEMENT LEARNING
  JOHN LANGFORD'S RESEARCH (INTERACTIVE LEARNING)

- LOW-CODE DEVELOPMENT OF AUTONOMOUS SYSTEMS
  PROJECT BONSAI

- SIX LEVELS OF AUTONOMY IN SELF-DRIVING CARS
  DEFINED BY THE SOCIETY OF AUTOMOTIVE ENGINEERS

- VALIDATION OF AUTONOMOUS SYSTEMS
  TECHNIQUES & TOOLS

- OTHER APPLICATIONS FOR AUTONOMOUS SYSTEMS
  e.g. SPACE EXPLORATION

MY REFERENCES

Essential Pre Read

Illustrated Guide to Artificial Intelligence : thoughtworks.com

What is Autonomy?

Lecture by David Lane: youtube.com/watch?v=oFNMk6JqfII
Blackberry: Ultimate guide to autonomous systems blackberry.qnx.com
WIRED Guide to Robots wired.com
Autonomous Systems with Microsoft AI: https://www.microsoft.com/
Beware The Automation Paradox : forrester.com/blogs

Papers/Engineering approach

Lloyd's Register Foundation's paper: Foresight review of robotics and autonomous systems ResearchGate 315787442
Towards A Holistic Software Systems Engineering Approach for Dependable Autonomous Systems: dl.acm.org

Machine teaching:

The next extension of machine learning intheblack.cpaaustralia.com.au
How people's expertise makes AI even more powerful blogs.microsoft.com

Reinforcement learning

Reinforcement learning (Research collection) at microsoft.com

John Langford's Research On Interactive Learning : arxiv.org/abs/2106.04887

Microsoft's cutting-edge machine-learning tool moves from the lab to the mainstream techrepublic.com

Testing Autonomy

Validation of Autonomous Systems infoq.com
Challenges and Current State-of-the-Art : researchgate.net 308092368
Towards a framework for certification of reliable autonomous systems: Springer.com
Verification and validation: liverpool.ac.uk

MY REFERENCES

Online Debates

Trust in Autonomous Systems (Gopal Ramchurn, Jack Stilgoe, Kate Devlin) : CogX youtube.com

Autonomous Systems Failures: Who is Legally and Morally Responsible? (Todd Murphey, Ryan Calo, Madeleine Clare Elish) : Northwestern Engineering youtube.com

Ethics, Morals

Autonomy in Moral and Political Philosophy plato.stanford.edu
Who Is Responsible When Autonomous Systems Fail? Cigionline.org
Ethics in action ethicsinaction.ieee.org

War

Modern Conflict and Artificial intelligence: cigionline.org
Lethal Autonomous Weapon systems: futureoflife.org
Dual Use technology: wikipedia.org
The Use Of Artificial Intelligence Technologies In Information And Psychological Warfare : researchgate.net 334454525

Standards, Regulation

International Legal Regulation of Autonomous Technologies: Cigionline.org

The IEEE Global Initiative on ethics: standards.ieee.org

Regulating Autonomous Systems: Beyond Standards: andrew.cmu.edu

Towards a framework for certification of reliable autonomous systems: link.springer.com

Ethics, regulation and the new artificial intelligence, part i: accountability and power: tandfonline.com

AI is mostly governed by ‘soft law’. But that is set to change: techmonitor.ai

The Laws and Regulation of AI and Autonomous Systems: ResearchGate.net 347376505